Abstract

Using Riboswitches to Uncover Hidden Biological Processes

and Novel Antibacterial Agents

Ruben Morais Atilho

2019

Riboswitches are noncoding RNAs (ncRNAs) that regulate gene expression in response to the binding of specific small molecule or ion ligands. Of the 40 riboswitch classes that have been experimentally validated to date, many control fundamental metabolic processes such as the biosynthesis and transport of amino acids, enzyme cofactors, and nucleotides. Based on the abundances and distributions of these known riboswitch classes, a large number of additional riboswitch classes are likely yet to be discovered. Unfortunately, the vast majority of these riboswitch classes are probably exceedingly rare or respond to ligands whose roles in biology were poorly known until recently. Examples of exceedingly rare riboswitch classes that were hidden due to their mistaken classification as members of known riboswitch classes is described in Chapter 2. I demonstrate that rare variants of the FMN riboswitch class exhibit altered ligand specificity and are likely used to sense flavin degradation products to activate transporters for their detoxification. Additionally, Chapter 3 details how my colleagues and I determined that subtype 1 *ykkC* motif RNAs selectively bind guanidine, a compound which was never known to be broadly important in bacteria until this riboswitch class was experimentally validated. I show that guanidine selectively induces reporter gene expression and the multidrug efflux pumps commonly controlled by this RNA likely function as guanidine transporters.

Given the rarity of the remaining riboswitch classes, current bioinformatic search techniques that rely solely on comparative sequence analysis are likely reaching their limit for differentiating riboswitches from false positives. In Chapter 4, I describe an improved computational search pipeline that identifies intergenic regions most likely to serve as templates for the production of structured ncRNAs. I demonstrate that a prominent riboswitch candidate discovered using this technique called the '*thiS* motif', function as sensors for the thiamin precursor HMP-PP. Furthermore, these findings provide another opportunity to learn how simple RNA sequences and structures could have selectively bound important nucleotide-derived metabolites in primitive organisms of the RNA World.

In addition to exploring the functional capabilities of RNA and providing opportunities to expose unique facets of biology, it is also possible to utilize riboswitches as tools to identify molecules as leads for the development of new antibacterial agents. In Chapter 5, I demonstrate how my colleagues and I used ZTP riboswitches as RNA-based small molecule sensors in a high-throughput screen to identify compounds that adversely affect folate metabolism in bacteria. We demonstrate that our assay can also be used to study structure-activity relationships of antifolate derivatives which can be used to develop compounds with improved activity.

Using Riboswitches to Uncover Hidden Biological Processes

and Novel Antibacterial Agents

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

by

Ruben Morais Atilho

Dissertation Director: Ronald R. Breaker

May 2019

Table of Contents

List of Figures

List of Tables

Acknowledgements

First and foremost, I want to thank my adviser Ronald Breaker for all his years of unwavering support and mentorship. I am tremendously grateful to Ron for everything he has taught me. I will never forget his contagious enthusiasm for knowledge, and impeccable work ethic, which I will always try to emulate.

I would like to thank Scott Strobel and Matt Simon, members of my thesis committee, for their time, interest and guidance. I am extremely grateful for their letters of recommendation and insightful comments which helped me obtain the National Science Foundation Graduate Research Fellowship. I would also like to acknowledge Wade Winkler at The University of Maryland for serving as my outside reader and providing valuable comments on my thesis.

I would like to thank all the members of the Breaker Lab who were incredibly supportive and who directly contributed to the research presented in this thesis. A huge thanks to both Adam Roth and Narasimhan Sudarsan whose immense knowledge and expertise has helped me on numerous occasions. A special thanks to Etienne Greenlee, Kevin Perkins, Sarah Malkowski, and Cynthia Megyola, who I am extremely grateful for their friendship and advice over the years.

I would like to thank Carol Fenn and Nicolas Carrasco from Quinnipiac University, for giving me the courage to apply for graduate school and the strong educational foundation needed to be successful.

I am grateful to all my friends and family, who have wished me luck along the way and kept me going through the tough times. Thank you Bigyan Dahal, Dipesh Khanal, Joelvito Villalus, Matt Ballas, Karthik Yarlagadda, Evan Koumbaros, Chris Becker,

Craig Richard, Dan Lavigne, Kevin Perkins, Etienne Greenlee, Janelle Lora, Larry Bowman, Tyler Smith, Danilo Rojas-Velasquez, and Juan Guzman. My Aunt Minda, Uncle Tavio, and cousins Jason and Colin, who always welcome me into their home with open arms. My parents and my brother Steven, for all their continual support and love throughout my years, always believing in me, and teaching me to never give up.

Last but certainly not the least, I would like to thank my fiancé Kirsten Knecht. I feel tremendously blessed for all these years we have spent together, first at Quinnipiac University getting our Bachelors and now at Yale University getting our PhDs. I cannot wait to see what the future holds for us. As long as we are by each other's side, I feel like anything is possible.

Chapter One

Introduction

Riboswitches Regulate Gene Expression

Riboswitches are structured noncoding RNAs typically found in the 5′ untranslated regions (UTR) of mRNAs (Breaker, 2011; Serganov and Nudler, 2013; Sherwood and Henkin, 2016). By selectively binding a small molecule or ion ligand, riboswitches are able to sense changes in ligand concentration and tune the expression of downstream open reading frames (ORFs). The architecture of these regulatory RNAs typically comprise of a highly conserved metabolite-binding aptamer domain and an expression platform. Conformational changes within the aptamer domain brought about by ligand binding causes the expression platform to activate or repress the expression of downstream ORFs. While the most common expression platforms control gene expression via intrinsic transcription termination or translation initiation, many additional expression platform types have been discovered (Breaker, 2012). In this chapter, the various types of ligands sensed by riboswitches, strategies used to discover them, and potential biotechnological applications are discussed.

Ligands for Known Riboswitch Classes

Approximately 40 riboswitch classes have been validated to date that sense fundamental metabolites, ions, and signaling compounds to regulate gene expression in various bacterial species (McCown et al., 2017; Serganov and Nudler, 2013; Sherwood and Henkin, 2016). Interestingly, an unusually large number of these riboswitch classes sense ligands that are derived from RNA nucleotides hinting they might be ancient in origin (Breaker, 2010; Breaker 2012; Nelson and Breaker, 2017). Perhaps RNA world organisms once used similar RNAs to regulate very complex metabolic states before the

emergence of proteins. Below is a comprehensive summary of the different types of ligands sensed by riboswitches.

Coenzymes. The most common ligand bound by riboswitches are coenzymes which are small organic molecules that are required by certain enzymes to carry out catalysis (McCown et al. 2017). Currently the nine coenzymes that have been validated as riboswitch ligands include flavin mononucleotide (FMN) (Mironov et al., 2002; Winkler et al. 2002a), adenosylcobalamin (AdoCbl) (Nahvi et al., 2002), aquacobalamin (AqCbl) (Johnson et al., 2012), thiamin pyrophosphate (TPP) (Winkler et al., 2002), S-adenosylmethionine (SAM) (Epshtein et al. 2003; McDaniel et al. 2003; Winkler et al. 2003), S-adenosylhomocysteine (SAH) (Wang et al., 2008), tetrahydrofolate (THF) (Ames et al., 2010), and molybdenum (MoCo) and tungsten cofactor (Wco) (Regulski et al., 2008). These coenzymes, which are largely derived from RNA nucleotides or their precursors, have been proposed to be present during the RNA world before proteins had emerged (White III, 1976). If true, riboswitch aptamers for these ancient molecules might represent modern version of ancient RNAs that once utilized these coenzymes long ago. The question then arises, where are the riboswitches that sense the numerous other fundamental coenzymes such as pyridoxal phosphate (PLP), coenzyme A, and nicotinamide adenine dinucleotide (NAD)? It could be very possible that there are more coenzyme-sensing riboswitch classes that have yet to be discovered.

Nucleotide Derivatives. Additional RNA derived ligands recognized by riboswitches include guanine (Mandal et al., 2003), adenine (Mandal and Breaker, 2004), 2′-

deoxyguanosine (2′-dG) (Kim et al. 2007; Weinberg et al. 2017), prequeuosine$_1$ (preQ$_1$) (Roth et al., 2007; Meyer et al., 2008; McCown et al. 2014), PRPP, (d)NDPs (Sherlock et al., 2019), and HMP-PP (Atilho et al., in press). The conserved sequence and secondary structure of guanine, adenine and 2′-dG are closely related but due to key mutations of highly-conserved nucleotides within the aptamer binding pocket they are able to selectively recognize different ligands. Members of another purine-sensing class are among the smallest of all known riboswitches and recognize the guanine-derived nucleobase preQ$_1$, which is known to be incorporated in the wobble position of tRNAs containing the GUN anticodon (encoding Tyr, His, Asp and Asn) and then further modified to yield queuosine (Q) (Iwata-Reuyl, 2008). Riboswitches that sense PRPP (Sherlock et al., 2018) and (d)NDP (Sherlock et al., 2019) were identified from a subset of *ykkC* motif RNAs. While many *ykkC* RNAs selectively recognize guanidine, 30% of all RNAs contain mutations in the binding pocket and have changed ligand specificity. Recently, a riboswitch class that senses the thiamin pyrimidine precursor HMP-PP (Atilho et al., in press) was discovered to control genes involved in the biosynthesis of the thiamin thaizole precursor HET-P. More information about HMP-PP riboswitches can be found in Chapter 4.

Signaling Molecules. Another striking trend among ligands sensed by riboswitches is the abundance of RNA-based signaling molecules. The signaling molecules sensed by riboswitches include c-di-GMP (Sudarsan et al., 2008; Lee et al., 2010), c-di-AMP (Nelson et al. 2013), c-AMP-GMP (Kellenberger et al., 2015; Nelson et al., 2015), ZTP (Kim et al. 2015), and ppGpp (Sherlock et al., 2018). Similar to coenzymes, the

connection between riboswitches and signaling molecules formed from ribonucleotide components again suggest that these compounds might have once been used in the RNA world long before proteins first emerged (Nelson and Breaker, 2017).

Ions. Currently, ion-binding riboswitches that recognize Mg^{2+} (Dann et al., 2007, Cromie et al., 2006), Mn^{+2} (Dambach et al., 2015, Price et al., 2015), Ni^{+2} and Co^{+2} (Furukawa et al., 2015), and F^{-} (Baker et al., 2012) have been discovered to date. Given the importance of inorganic ions in biology, it is very likely that additional riboswitch classes exist for metal ions. However, due to the inherent ability for polyanionic RNA polymers to interact with cationic ligands, extensive amount of evidence will be necessary to convincingly prove any new ion-responsive riboswitch claims.

Amino Acids. Of the twenty common amino acids used in biology, surprisingly, only three are sensed by riboswitch aptamers. The amino acids directly sensed by riboswitches include lysine (Grundy et al., 2003; Sudarsan et al., 2003b), glycine (Mandal et al., 2004), and glutamine (Ames and Breaker, 2011; Klähn et al., 2018). Perhaps the vast majority of biosynthesis and transport gene of most amino acids are regulated by T-box regulatory RNA elements (Green et al., 2010) and attenuator systems (Vitreschak et al., 2004) that indirectly sense amino acid deficiency. Given the metabolic importance of genes controlled by many riboswitch classes, such as lysine amino acid biosynthesis, studies have shown that riboswitches have the potential to serve as new classes of antibacterial drug targets (Blount et al., 2007; Blount and Breaker, 2006).

Other Metabolites. The last three ligands sensed by riboswitches do not fall into any of the ligand categories mention above and each regulate a different aspect of bacterial biology. Members of the first riboswitch class sense the modified sugar compound glucosamine-6-phosphate (GlcN6P) and utilize a metabolite-cofactor-dependent ribozyme as part of their mechanism of genetic control (Winkler et al., 2004). The glmS gene encodes the first step in the pathway for the production of UDP-GlcNAc, which is subsequently used in the process of cell wall biosynthesis (Milewski et al., 2002). The second riboswitch class responses to a diverse collection of azaaromatic compounds that contain nitrogen-containing aromatic heterocycles (Le et al., 2016). Members of this riboswitch class associate primarily with a single gene, called *yjdF*, whose protein function is currently unknown. It has been proposed that this riboswitch class might serve as part of a genetic response system to deal with toxic chemical similar to azaaromatics. The final ligand in this miscellaneous category that is sensed by riboswitches is guanidine. Currently, three distinct riboswitch classes called guanidine-I, -II, and –III have been validated to selectively bind guanidine and regulate genes used to overcome guanidine toxicity in bacteria (Nelson et al., 2017; Sherlock and Breaker, 2017; Sherlock et al., 2017). More information about guanidine-I riboswitch class can be found in Chapter 3.

Riboswitch Discovery Methods

The continued discovery of structured non-coding RNAs in bacteria will provide insights into the expanding roles and capabilities of RNA molecules and their potential to serve as targets and tools for the development of new antibacterial agents. Below is a summary of

the various riboswitch discovery methods that ultimately had a riboswitch candidate biochemically and/or genetically validated.

Literature-based Riboswitch Discovery. Prior to the discovery of riboswitches, gene control systems in bacteria were largely restricted to proteins factors that modulate gene expression in response to changing cellular conditions. However, as examples of engineered RNAs forming highly selective aptamers (Famulok, 1999; Hermann and Patel, 2000) and aptamer-ribozyme molecular switches (Soukup and Breaker, 2000; Seetharaman et al., 2001) were shown to be possible, several researchers started speculating whether mRNA could directly regulate gene expression without the need of protein factors (Gold et al., 1997; Stormo and Ji, 2001).

After the initial discovery that *btuB* mRNA leaders serve as metabolite-sensing genetic switches (Nahvi et al., 2002), additional examples of mRNA leaders, such as the *RFN* element (Gelfand et al., 1999) and the *thi* box (Miranda-rios et al., 2001), were identified from the literature. Similar to *btuB* mRNA leader sequences, these RNAs also were found to be important for genetic regulation of the downstream ORF but no protein factor had yet been identified. This approach allowed for the discovery of multiple riboswitch classes including classes that sense adenosylcobalamin (AdoCbl) (Nahvi et al., 2002), Thiamine pyrophosphate (TPP) (Winkler et al., 2002), flavin mononucleotide (FMN) (Mironov et al., 2002; Winkler et al. 2002a), S-adenosylmethionine (SAM) (Epshtein et al. 2003; McDaniel et al. 2003; Winkler et al. 2003), guanine (Mandal et al., 2003), lysine (Grundy et al. 2003; Sudarsan et al. 2003b), and MoCo (Regulski et al., 2008). However, because the regulation of many genes in Eubacteria are not widely

understood and genetic-based approaches being extremely time consuming, computational-based approaches were created to discover the remaining highly-conserved metabolite-binding RNA motifs that exist in bacteria.

Comparative Sequence Analysis. Bioinformatics-based searches (Barrick et al., 2004; Weinberg et al., 2007; Weinberg et al., 2010; Weinberg et al., 2017a) have been incredibly successful in identifying new riboswitch classes. These search strategies rely on comparative sequence analysis to identify new riboswitch candidates by analyzing the 5' UTRs of genes for highly conserved sequence and secondary-structure elements that are used to form the ligand binding riboswitch aptamer domain. Sequence and secondary structure models are created by clustering IGRs by sequence similarly based on BLAST homology comparisons (Altschul et al., 1997) and then inferring a conserved secondary structure within each group using CMfinder (Yao et al., 2006).

One caveat of comparative sequence analysis is that an input IGR sequence is needed in order to run the homology and secondary structure prediction programs. With over 37 billion base pairs present within bacterial genomic sequence datasets (Weinberg et al., 2017a), the computational time and memory needed to analyze every IGR at once makes it essentially not feasible. Instead, a subsets of IGRs must be selected that are enriched for riboswitch candidates by identify IGRs that contain architectural features of known riboswitches aptamers. Architectural features of known riboswitch used to select IGR subsets include, 1) riboswitches typically reside in unusually large IGRs, 2) riboswitches from certain classes tend to be enriched in related bacterial lineages, 3) riboswitches typically utilize transcription terminators to control gene expression, 4) riboswitches tend

to cluster upstream of certain gene classes, and 5) riboswitch tend to have roughly equal distribution of all four nucleotide types.

Computational search strategies can also couple multiple parameters to enrich for structured ncRNAs (Meyers et al., 2009; Stav et al., in press). Riboswitches often reside in relatively long IGRs that contain equal distribution of all four nucleotide types. This is because of their need to form complex and stable secondary and tertiary structures that can selectively recognize a metabolite or ion ligand. IGRs from an individual sequence genome can therefore be sorted by these two parameters to identify IGRs that cluster with known structured ncRNAs. Once a candidate ncRNA is identified, comparative sequence analysis can be used to expand the number of natural representatives of each motif. To date, this approach has identified two novel riboswitch classes that selectively respond to the coenzyme *S*-adenosylmethionine (SAM) (Poiata et al., 2009) and the thiamin precursor HMP-PP (Atilho et al., in press) which is described in further detail in Chapter 4.

Variant Riboswitch Discovery. Riboswitch variants are RNAs wherein the consensus sequence and secondary-structure models highly resemble a previously-validated parent riboswitch class, but carry key mutations that have altered their ligand specificities. Unfortunately, bioinformatics algorithms have consistently clustered these variant RNAs amongst their more prominent riboswitch class until mutations in the ligand-binding core or changes in gene associations were noticed. These include variants of *ykkC* motif RNAs, FMN, purine, cobalamin, MoCo, and c-di-GMP-I riboswitch classes. Recently, a bioinformatics pipeline was constructed to systematically analyze known riboswitch

classes to find additional variant classes (Weinberg et al., 2017b). Using atomic-resolution structural information of the ligand binding site and riboswitch gene associations, several unique variants that carry unusual nucleotides within the ligand-binding site were identified. Specifically, a second riboswitch class for the nucleoside 2′-deoxyguanosine was discovered within metagenomic sequence datasets.

Riboswitch Variants

Similar to how certain groups of protein enzymes and receptors can evolve to change ligand specific, several examples of riboswitch variants have been reported wherein the ligand-biding aptamer domain has been mutated to selectively accommodate a different ligand. Below is a comprehensive summary of all the riboswitch classes where ligand specificity changes are known. The continued identification of variant riboswitch classes will allow us to better understand how small changes in RNA sequence can lead to minor or sometimes major changes in ligand specificity.

Purine Riboswitch Variants. Evidence that variant riboswitch classes can accrue mutations to change their ligand specificity was first revealed with purine-sensing riboswitches. Although guanine repressing the expression of purine biosynthesis and transport genes in *Bacillus subtilis* was known for quite some time (Christiansen et al., 1997), no regulatory protein factors that sense guanine to control gene expression was ever identified. Eventually, highly conserved aptamers located in the 5’ UTRs of purine metabolism genes in various bacterial species was shown to selectively bind guanine and control gene expression in the absence of proteins (Mandal et al., 2003).

Shortly after the discovery of guanine riboswitches, a subset of these RNAs were found to recognize adenine (Mandal and Breaker, 2004b) and control genes involved with the biosynthesis and transport of adenine. Despite having very similar sequence and secondary structures, each representative of the adenine subclass was shown to carry a single C-to-U mutation in the conserved ligand binding site within the aptamer. X-ray crystal structures of both guanine and adenine aptamers have subsequently confirmed that ligand specificity is due to Watson-Crick base pairing of the bound guanine and adenine ligands with cytidine and uridine, respectively (Batey et al., 2004; Serganov et al., 2004). In addition to this specific base pairing interaction, three additional uridine residues in both aptamers are involved in making direct hydrogen bond interactions to the other functional groups of the purine ligand.

Additional examples of rare variant riboswitch classes of guanine riboswitches include two classes of riboswitches that sense 2′-deoxyguanosine (2′-dG). To date, the 2′-dG-I riboswitch class has only been identified in *Mesoplasma florum* (Kim et al., 2007). Similarly, the recently discovered 2′-dG-II riboswitch class has only been found within metagenomic sequence data sets (Weinberg et al., 2017b). Both were identified using a bioinformatic search strategy to discover variants of currently known riboswitch classes by searching for sequences that deviate in key ligand binding positions. The consensus and secondary structure model of both riboswitch class resemble guanine riboswitches, but contain mutations located in the vicinity of the N9 positon of guanine. 2′-dG-I riboswitch are found upstream of genes encoding ribonucleotide reductase (Kim et al., 2007) which convert ribonucleotides into their deoxyribonucleotide counterparts. X-ray crystal structures of 2′-dG-I riboswitches from *Mesoplasma florum* have confirmed that

2′-dG-I riboswitches utilize several critical changes that rearrange the ligand binding site to accommodate the additional sugar moiety of 2′-deoxyguanosine (Pikovskava et al., 2011).

***ykkC* Motif RNA Variants.** Originally discovered using comparative sequence analysis to identify riboswitches over a decade ago (Barrick et al., 2004), *ykkC* motif RNAs nicely highlight how metabolite-binding RNAs can recognize a variety of ligands by simply mutating a few key residues within a general RNA aptamer scaffold. Currently, of the five different variant riboswitch classes present in *ykkC* motif RNAs, subtype 1 RNAs recognize guanidine (Nelson et al., 2017), subtype 2a RNAs recognize ppGpp (Sherlock et al., 2018a), subtype 2b RNAs recognize PRPP (Sherlock et al., 2018b), subtype 2c RNAs recognize nucleoside diphosphate compounds (Sherlock et al., 2019), and the target ligand for subtype 2d RNAs is currently not known.

Genes predicted to be under control of subtype 1 *ykkC* motif RNAs included urea carboxylases and their associated proteins, allophanate hydrolases, arginases, creatininases, nitrate/sulfate/bicarbonate transporters, and various multi-drug efflux pumps (Barrick et al., 2004; Meyer et al., 2011). Due to the eclectic distribution of metabolic genes and transporters it was hypothesized that subtype 1 *ykkC* RNAs might sense a toxic ligand. Eventually guanidine, a common protein denaturant and chaotropic agent, was discovered to be the natural ligand for this class of RNAs (Nelson et al., 2017). More information about the validation of guanidine-I riboswitches can be found in Chapter 3. High-resolution X-ray crystal structures have revealed that a selective binding pocket is formed by recognizing every functional group of guanidinium through

hydrogen bonds to guanine bases and phosphate oxygens (Reiss et al., 2017; Battaglia et al., 2017). Additionally, binding is further stabilized through cation-π interactions.

The ppGpp riboswitch class, previously called subtype 2a *ykkC* motif RNAs, are predicted to regulate genes involved in the biosynthesis and transport of branch-chain amino acids (BCAA), glutamine synthases, and ABC transporters (Sherlock et al., 2018a). Similar to the guanidine-I riboswitch class, ppGpp riboswitches are "ON-switches" that trigger gene expression when ppGpp is bound the riboswitch aptamer. These findings are consistent with the fact that BCAA biosynthesis is activated during the stringent response in Firmicutes (Wolz et al., 2010). During times of starvation, such as amino acid depletion, the alarmone ppGpp signals the stringent response which allows cells to divert resources away from growth and division and toward amino acid synthesis in order to promote survival until nutrient conditions improve. Unlike guanidine-I riboswitches, ppGpp riboswitches contain an additional helical element that joins the ends of the ligand-sensing domains and creates a tunnel for the direct and Mg^{2+}-mediated binding of ppGpp (Knappenberger et al., 2018; Peselis and Serganov, 2018).

The PRPP (subtype 2b) riboswitch class are commonly associated with genes for de novo purine biosynthesis (Sherlock et al., 2018b) and are typically found in a tandem arrangement with a guanine aptamer (Sudarsan et al., 2006). When PRPP accumulates in cells, genes involved in the production of the purine branch point intermediate inosine monophosphate (IMP) are activated. During the stringent response, when cells need to keep GTP levels low, PRPP riboswitches only promote AMP synthesis. Similar to ppGpp riboswitches, the riboswitch class that senses PRPP forms an additional P0 stem in order to extend the binding pocket of this larger ligand (Knappenberger et al., 2018; Peselis and

Serganov, 2018). Despite having a similar consensus sequence (Sherlock et al., 2018b), a single highly conserved guanine residue in subtype 2b RNAs differentiates PRPP and ppGpp aptamers (Knappenberger et al., 2018).

Recently, representatives of *ykkC* subtype 2c RNAs have been shown to recognize various adenosine and cytidine 5′-diphosphate molecules in either their ribose or deoxyribose forms (ADP, dADP, CDP, and dCDP) (Sherlock et al., 2019). However, due to the broad specificity of NUDIX hydrolases (McLennan, 2006), the gene most commonly associated with subtype 2c RNAs, it was not possible to definitively determine the true ligand of this variant class. Interestingly, among all the known *ykkC* RNA variant classes, subtype 2c RNAs most closely resemble PRPP riboswitches and contain tail regions which might form a P0 stem similar to PRPP riboswitches.

Currently, subtype 2d RNAs are the oldest orphan riboswitch candidate known to date (Sherlock et al., 2019). These RNAs closely resemble the consensus model for guanidine-I riboswitches such that that they also lack the structured tail regions at the 5' and 3' ends of the aptamer that form part of the sugar-phosphate binding pocket of subtypes 2a and 2b (Sherlock et al., 2019, Knappenberger et al., 2018). Additionally, subtype 2d RNAs are predicated to control a transporter of unknown function associated with phosphonate utilization (Sherlock et al., 2018b) and, therefore, might sense a small molecule or ion similarly to guanidine.

Cobalamin Riboswitch Variants. Previously known as "B_{12} box" RNAs, the *btuB* mRNA leader from *Escherichia coli* was the first validated riboswitch shown to regulate gene expression independent of any protein factor by directly sensing changing

concentrations of a vitamin B_{12} derivative called adenosylcobalamin (AdoCbl) (Nahvi et al., 2002). Cobalamin riboswitches are a widespread riboswitch class that typically regulate the biosynthesis and transport of cobalamin derivatives in bacteria. X-ray crystal structures of Cbl riboswitches (Johnson et al., 2012; Peselis et al., 2012) have shown that while most of the conserved nucleotides of the aptamer from the ligand binding site for the coenzyme, ligand recognition is primarily achieved through van der Waals interactions with various moieties of the corrin ring and only utilize two hydrogen bonds between the adenosyl moiety of AdoCbl and an adenosine of the aptamer J6-3 region. Variant riboswitches with altered ligand specificity were subsequently discovered that do not extensively alter binding site nucleotides to modify their ligand-binding specificity but utilize conformational differences in the J6-3 region to cause steric clashes with the adenosyl moiety of the coenzyme (Johnson et al., 2012). Representatives of this variant riboswitch class were found to recognize aquacobalamin (AqCbl), a cobalamin derivative wherein the adenosyl moiety of the coenzyme is replaced by water, with 500-fold higher affinity than AdoCbl. Additional variant classes hidden within known riboswitches could perhaps use similar approaches of altering ligand specificity by utilizing conformational difference of nucleotides within of the ligand-binding site. However, current bioinformatic search approaches that focus on mutational changes in the binding site are likely miss on these examples (Weinberg et al., 2017b).

MoCo Riboswitch Variants. Genes encoding molybdenum cofactor (Moco) biosynthesis enzymes and proteins that utilize Moco as a coenzyme are commonly regulated by a highly conserved RNA motif that forms a five stem (P1-P5) junction with

a GNRA tetraloop (Regulski et al., 2008). In *E.coli*, transcription of the molybdopterin biosynthesis operon is upregulated by anaerobic protein transcription factors (Spiro and Guest, 1990; Anderson et al., 2000), however, if sufficient levels of Moco is already present, translation of this operon is prevented when the Moco RNA aptamer in the mRNA leader binds Moco. Organisms found not to require molybdenum have instead evolved to utilize tungsten, a transition metal located in the same column of the periodic table, and the analogous coenzyme Wco in its place (Johnson et al., 1993). In these tungsten utilizing organism, a variant class of Moco riboswitches was discovered to regulate Wco metabolism genes (Regulski et al., 2008). The consensus and secondary structure of Wco riboswitches look identical to Moco riboswitch class with the exception of a missing P3 stem. Due to the instability of Moco and Wco, structural and biochemical analysis of these RNAs have not been possible. It is likely that these two riboswitch classes utilize distinctive structural features that would allow the binding pockets to discriminate between these two nearly identical cofactors.

c-di-GMP-I Riboswitch Variants. Two structurally different riboswitch classes exist that selectivity recognize c-di-GMP to control genes involved in pilus assembly, motility, chemotaxis sensing and pathogenesis (Sudarsan et al., 2008; Lee et al., 2010). X-ray crystal structures of c-di-GMP-I riboswitches show that the ligand binding site lies within the three-helix junction and uses two conserved nucleotides (G20 and C92) within the aptamer and base-stacking interactions to bind c-di-GMP (Smith et al., 2009; Kulshina et al., 2009; Smith and Strobel, 2011; Shanahan et al., 2011). In search for additional riboswitch classes of natural cyclic dinucleotides, variants of the c-di-GMP-I

riboswitches class that carry a G20A mutation were discovered to selectively bind c-AMP-GMP in various strains of Deltraproteobacteria (Nelson et al., 2015; Kellenberger et al., 2015). These variant RNAs are typically found upstream of genes not commonly controlled by c-di-GMP-I riboswitches, such as genes involved in exoelectrogenesis, pilin formation and aromatic amino acid biosynthesis. Structural analysis of c-AMP-GMP riboswitches show that in addition to the G20A mutations, their P2 and P3 stems fold into a unique orientation depending on which cyclic dinucleotide binds. However, promiscuous c-di-GMP-I riboswitches that have their G20 mutated to an adenine adopt nearly identical stem orientations (Ren et al., 2015).

FMN Riboswitch Variants. Members of the riboswitch class that bind to flavin mononucleotide (FMN) usually control the expression of genes required for the biosynthesis and transport of the coenzyme precursor riboflavin (Gelfand et al., 1999; Mironov et al., 2002; Winkler et al., 2002). Examples of FMN riboswitch variants, called CD3299 RNAs, were previously reported (Blount et al. 2012; Blount et al., 2013) to exist in several strains of *Clostridium difficile*. Examples of this variant riboswitch class never associate with known FMN metabolism genes, and differ from the FMN riboswitch consensus sequence at highly conserved nucleotides that are known to form direct contacts with FMN. Additional examples were subsequently uncovered in a bioinformatics study that purposefully examined known riboswitch classes for the existence of variant riboswitch aptamers that have possibly altered their ligand binding specificities (Weinberg et al., 2017b). In Chapter 2, I show evidence for two highly similar FMN variant subtypes, and demonstrate that both subtypes completely reject

FMN (Atilho et al., 2019). Additionally, I show that subtype 2 RNAs strongly bind riboflavin and certain light-mediated FMN degradation products. We hypothesize that some variant FMN RNAs might have evolved to recognize flavin derivatives to activate transporters involved in their disposal.

Riboswitches as Antibacterial Drug Targets

Since their initial discovery in 1929, antibiotics have become an important tool for controlling bacterial infections (Fleming, 1929; Demain and Sanchez, 2009). However, due to the increasing number of drug-resistant bacteria, many of our antibacterial agents are becoming less effective (Nikaido, 2009; Fischbach and Walsh, 2009). As a consequence, multidrug-resistant bacteria are a major threat to human health (Talbot et al., 2006; Koul et al., 2011). To address this problem, new chemical scaffolds are needed to reinvigorate antibacterial drug discovery. Targeting bacterial classes of riboswitches might help restock our diminishing arsenal of antibiotics. Because riboswitches are not found in human, novel compounds can be developed that target riboswitches and disrupt the regulation of vital genes in bacteria (Blount and Breaker, 2006). Below is a summary of the several known antibacterial compounds that function by targeting riboswitches.

Analogs That Target TPP Riboswitches. Pyrithiamine (PTPP), once phosphorylated inside cells, is chemically similar to TPP but contains a pyridinium ring instead of a central thiazole ring. Before there was any evidence of PTPP targeting TPP riboswitches (Sudarsan et al., 2005), it was known for quite some time that pyrithiamine was able to inhibit the growth of several bacterial and fungal species (Robbins, 1941; Wooley and

White, 1943). Furthermore, bacterial strains resistant to pyrithiamine were discovered to have mutations in conserved regions of TPP riboswitches which likely confer resistance though derepression of thiamine biosynthesis.

Analogs That Target Lysine Riboswitches. Analogs of lysine were also previously known to inhibit bacterial growth (Shiota et al., 1958; McCord et al., 1957). Both L-aminoethylcysteine (AEC) and DL-4-oxalysine, are now known to bind lysine riboswitches and repress lysine biosynthesis in vivo (Sudarsan et al., 2003; Blount et al., 2007). Bacterial strains resistant to these analogs have also been shown to contain mutations that map to the lysine aptamer domain (Vold et al. 1975; Lu et al. 1992; Patte et al. 1998). Similar to the effects of pyrithiamine, these mutations likely confer resistance through derepression of lysine biosynthesis.

Analogs That Target FMN Riboswitches. Roseoflavin is a chemical analog of FMN in which a methyl group in the flavin ring is replaced by a dimethylamino group. While the antimicrobial properties of roseoflavin were known for quite some time (Otani et al., 1974), the exact mechanism of action was unknown until recently (Lee et al., 2009). In-line probing assays show that roseoflavin binds to FMN riboswitches and dysregulates riboflavin production in vivo. Similar to pyrithiamine, roseoflavin most likely gets phosphorylated inside cells to form roseoflavin monophosphate (RoF-P). Roseoflavin resistant strains contain mutations in FMN riboswitches that result in the derepression of FMN biosynthesis (Kukanova et al., 1982). Recently, two independent studies have developed novel FMN riboswitch-binding antibacterial compounds that inhibit the

growth of *Clostridium difficile* (Blount et al., 2015; Vicens et al., 2018), and *E. coli, Staphylococcus aureus* (MRSA) and *Enterococcus faecalis* (Howe et al., 2015; Howe et al., 2016; Wang et al., 2017).

Riboswitches as Tools for Antibacterial Drug Discovery

While target-based screening, which involves screening for small-molecule protein inhibitors in vitro, has dominated modern drug discovery the past few decades, there is a renewed interest in whole-cell phenotypic screening due to problems associated with target-based approaches (Swinney and Anthony, 2011; Zheng et al., 2013). However, the main issue with phenotypic screens is target identification once a small-molecule lead is identified. One way to overcome this challenge is by utilizing riboswitches fused to reporter genes as a tool to identify compounds that perturbed certain metabolic processes in bacteria. Recently, a fluoride riboswitch reporter was used to identify compounds that would enhance fluoride toxicity (Nelson et al., 2015). In Chapter 5, my colleagues and I demonstrate how a ZTP-riboswitch reporter system can be used in a high-throughput screen to identify compounds that disrupt purine biosynthesis via folate inhibition in *E. coli* (Perkins et al., in press).

References

Altschul SF, Madden TL, Schaffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ. 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**: 3389-3402.

Ames TD, Breaker RR. 2011. Bacterial aptamers that selectively bind glutamine. *RNA Biol* **8**: 82–89.

Ames TD, Rodionov DA, Weinberg Z, Breaker RR. 2010. A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate. *Chem Biol.* **17**: 681-685.

Anderson LA, McNairn E, Lubke T, Pau RN, Boxer DH. 2000. ModE-dependent molybdate regulation of the molybdenum cofactor operon *moa*. *Escherichia coli*. *J Bacteriol* **182:** 7035–7043.

Atilho RM, Perkins KR, Breaker RR. 2019. Rare variants of the FMN riboswitch class in *Clostridium difficile* and other bacteria exhibit altered ligand specificity. *RNA*. **25**: 23-34.

Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, Breaker RR. 2012. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335**: 233–235.

Barrick JE, Corbino KA, Winkler WC, Nahvi A, Collins J, Lee M, Roth A, Sudarsan N, Jona I, Wickiser JK, Breaker RR. 2004. New RNA motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci USA* **101**: 6421-6426.

Batey RT, Gilbert SD, Montange RK. 2004. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**: 411–415.

Battaglia RA, Price IR, Ke A. 2017. Structural basis for guanidine sensing by the ykkC family of riboswitches. *RNA* **23**: 578−585.

Blount K F, Breaker RR. 2006. Riboswitches as antibacterial drug targets. *Nat. Biotechnol* **24**: 1558–1564.

Blount KF, Megyola C, Plummer M, Osterman D, O'Connell T, Aristoff P, Quinn C, Chrusciel A, Poel TJ, Schostarez HJ, Stewart CA, Walker DP, Wuts PG, Breaker RR. 2015. Novel riboswitch-binding flavin analog that protects mice against Clostridium difficile infection without inhibiting cecal flora. *Antimicrob Agents Chemother* **59**: 5736-5746.

Blount KF, Wang JX, Lim J, Sudarsan N, Breaker RR. 2007. Antibacterial lysine analogs that target lysine riboswitches. *Nat Chem Biol.* **3**: 44-49.

Breaker RR. 2010. RNA second messengers and riboswitches: relics from the RNA world? *Microbe* **5**: 13-20.

Breaker RR. 2011. Prospects for riboswitch discovery and analysis. *Mol Cell* **43**: 867–879.

Breaker RR. 2012. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol* **4**: 1-15.

Christiansen LC, Schou S, Nygaard P, Saxild HH. 1997. Xanthine metabolism in *Bacillus subtilis*: characterization of the *xpt–pbuX* operon and evidence for purine- and nitrogen-controlled expression of genes involved in xanthine salvage and catabolism. *J. Bacteriol*. **179**: 2540–2550.

Cromie MJ, Shi Y, Latifi T, Groisman EA. 2006. An RNA sensor for intracellular Mg^{2+}. *Cell* **125**: 71–84.

Dambach M, Sandoval M, Updegrove TB, Anantharaman V, Aravind L, Waters LS, Storz G. 2015. The ubiquitous yybP-ykoY riboswitch is a manganese-responsive regulatory element. *Mol Cell* **57**: 1099–1109.

Dann CE III, Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC. 2007. Structure and mechanism of a metal-sensing regulatory RNA. *Cell* **130**: 878–892.

Demain AL, Sanchez S. 2009. Microbial drug discovery: 80 years of progress. *J. Antibiot. (Tokyo)*. **62:** 5–16.

Epshtein V, Mironov AS, Nudler E. 2003. The riboswitch-mediated control of sulfur metabolism in bacteria. *Proc Natl Acad Sci USA* **100**: 5052–5056.

Famulok M. Oligonucleotide aptamers that recognize small molecules. 1999. *Curr Opin Struct Biol.* **9**: 324-329.

Fischbach MA, Walsh CT. 2009. Antibiotics for emerging pathogens. *Science* **325:** 1089–1093.

Fleming A. 1929. On the antibacterial action of cultures of Penicillium, with special reference to their use in the isolation of *B. influenzae*. *Br. J. Exp. Pathol.* **10**: 226–236.

Furukawa K, Ramesh A, Zhou Z, Weinberg Z, Vallery T, Winkler WC, Breaker RR. 2015. Bacterial riboswitches cooperatively bind Ni^{2+} or Co^{2+} ions and control expression of heavy metal transporters. *Mol Cell* **57**: 1088–1098.

Gelfand MS, Mironov AA, Jomantas J, Kozlov YI, Perumov DA. 1999. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet.* **15**: 439-442.

Gold L, Brown D, He Y, Shtatland T, Singer BS, Wu Y. 1997. From oligonucleotide shapes to genomic SELEX: Novel biological regulatory loops. *Proc. Natl. Acad. Sci. USA* **94**: 59–64.

Grundy FJ, Lehman SC, Henkin TM. 2003. The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc Natl Acad Sci* **100**: 12057–12062.

Hermann T, Patel DJ. 2000. Adaptive recognition by nucleic acid aptamers. *Science* **287**: 820–825.

Howe JA, Wang H, Fischmann TO, et al. 2015. Selective small-molecule inhibition of an RNA structural element. *Nature* **526**: 672-677.

Howe JA, Xiao L, Fischmann TO, Wang H, Tang H, Villafania A, Zhang R, Barbieri CM. 2016. Atomic resolution mechanistic studies of ribocil: A highly selective unnatural ligand mimic of the E. coli FMN riboswitch. *RNA Biol* **13**: 946-954.

Iwata-Reuyl D. 2008. An embarrassment of riches: the enzymology of RNA modification. *Curr. Opin. Chem. Biol*. **12**: 126–133.

Johnson JE, Jr, Reyes FE, Polaski JT, Batey RT. 2012. B12 cofactors directly stabilize an mRNA regulatory switch. *Nature* **492**:133–137.

Johnson JL, Rajagopalan KV, Mukund S, Adams MW. 1993. Identification of molybdopterin as the organic component of the tungsten cofactor in four enzymes from hyperthermophilic Archaea. *J Biol Chem* **268:** 4848–4852.

Kellenberger CA, Wilson SC, Hickey SF, Gonzalez TL, Su Y, Hallberg ZF, Brewer TF, Iavarone AT, Carlson HK, Hsieh YF, Hammond MC. 2015. GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. *Proc Natl Acad Sci USA* **112**: 5383-5388.

Kim PB, Nelson JW, Breaker RR. 2015. An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol Cell* **57**: 317–328.

Klähn S, Bolay P, Wright PR, Atilho RM, Brewer KI, Hagemann M, Breaker RR, Hess WR. 2018. A glutamine riboswitch is a key element for the regulation of glutamine synthetase in cyanobacteria. *Nucleic Acids Res.* **46**:10082-10094.

Knappenberger AJ, Reiss CW, Strobel SA. 2018. Structures of two aptamers with differing ligand specificity reveal ruggedness in the functional landscape of RNA. *Elife* **7**.

Koul A, Arnoult E, Lounis N, Guillemont J, Andries K. 2011. The challenge of new drug discovery for tuberculosis. *Nature* **469**: 483–490.

Kukanova A, Zhdanov VG, Stepanov AI. 1982. *Bacillus subtilis* mutants resistant to roseoflavin. *Genetika* **18**: 319-321.

Kulshina N, Baird NJ, Ferré-D'Amaré AR. 2009. Recognition of the bacterial second messenger cyclic diguanylate by its cognate riboswitch. *Nat Struct Mol Biol* **16**: 1212–1217.

Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR. 2010. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* **329**: 845–848.

Lee ER, Blount KF, Breaker RR. 2009. Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol* **6**:187-194.

Li S, Hwang XY, Stav S, Breaker RR. 2016. The yjdF riboswitch candidate regulates gene expression by binding diverse azaaromatic compounds. *RNA* **22**: 530-41.

Lu Y, Shevtchenko TN, Paulus H. 1992. Fine structure mapping of *cis*-acting control sites in the *lysC* operon of *Bacillus subtilis*. *FEMS Microbiol. Lett.* **92:** 23–27.

Mandal M, Boese B, Barrick JE, Winkler WC, Breaker RR. 2003. Riboswitches control fundamental biochemical pathways in Bacillus subtilis and other bacteria. *Cell* **113**: 577–586.

Mandal M, Breaker RR. 2004. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* **11**: 29–35.

Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR. 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**: 275–279.

McCord T, Ravel J, Skinner C, Shive W. 1957. DL-4-oxalysine, an inhibitory analog of lysine. *J. Am. Chem. Soc.* **79**: 5693–5696.

McCown PJ, Corbino KA, Stav S, Sherlock ME, Breaker RR. 2017. Riboswitch diversity and distribution. *RNA* **23**: 995-1011.

McCown PJ, Liang JJ, Weinberg Z, Breaker RR. 2014. Structural, functional, and taxonomic diversity of three preQ$_1$ riboswitches. *Chem Biol* **21**: 880–889.

McDaniel BAM, Grundy FJ, Artsimovitch I, Henkin TM. 2003. Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA. *Proc Natl Acad Sci* **100**: 3083–3088.

McLennan AG. 2006. The Nudix hydrolase superfamily. *Cell. Mol. Life Sci.* **63**: 123−143.

Meyer MM, Ames TD, Smith DP, Weinberg Z, Schwalbach MS, Giovannoni SJ, Breaker RR. 2009. Identification of candidate structured RNAs in the marine organism 'Candidatus Pelagibacter ubique'. *BMC Genomics* **10**: 268.

Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N, Breaker RR. 2011. Challenges of ligand identification for riboswitch candidates. *RNA Biol*. **8**: 5−10.

Meyer MM, Roth A, Chervin SM, Garcia GA, Breaker RR. 2008. Confirmation of a second natural preQ1 aptamer class in Streptococcaceae bacteria. *RNA* **14**: 685-695.

Milewski S. 2002. Glucosamine-6-phosphate synthase: the multi-facets enzyme. *Biochim. Biophys. Acta* **1597**: 173–192.

Miranda-ríos J, Navarro M, Soberón M. 2001. A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc Natl Acad Sci USA* **98**: 9736-9741.

Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E. 2002. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**: 747–756.

Nahvi AS, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR. 2002. Genetic control by a metabolite binding mRNA. *Chem Biol* **9**: 1043–1049.

Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB, Breaker RR. 2017. Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Mol Cell* **65**: 220–230.

Nelson JW, Breaker RR. 2017. The lost language of the RNA World. *Sci Signal* **10**: 483.

Nelson JW, Plummer MS, Blount KF, Ames TD, Breaker RR. 2015. Small molecule fluoride toxicity agonists. *Chem Biol* **22**: 527-534.

Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX, Breaker RR. 2013. Riboswitches in eubacteria that sense the second messenger c-di-AMP. *Nat Chem Biol* **9**: 834–839.

Nelson JW, Sudarsan N, Phillip GE, Stav S, Lünse CE, McCown PJ, Breaker RR. 2015. Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proc Natl Acad Sci* **112**: 5389–5394.

Nikaido H. 2009. Multidrug resistance in bacteria. *Annu. Rev. Biochem.* **78:** 119–146.

Otani S, Takatsu M, Nakano M, Kasai S, Miura R. 1974. Letter: Roseoflavin, a new antimicrobial pigment from Streptomyces. *J Antibiot (Tokyo)* **27**: 86-87.

Patte JC, Akrim M, Méjean V. 1998. The leader sequence of the *Escherichia coli lysC* gene is involved in the regulation of *LysC* synthesis. *FEMS Microbiol. Lett* **169:** 165–170.

Peselis A, Serganov A. 2012. Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. *Nat Struct Mol Biol*. **19**: 1182-1184.

Peselis A, Serganov A. 2018. ykkC riboswitches employ an add-on helix to adjust specificity for polyanionic ligands. *Nat Chem Biol.* **14**: 887-894.

Pikovskaya O, Polonskaia A, Patel DJ, Serganov A. 2011. Structural principles of nucleoside selectivity in a 2'-deoxyguanosine riboswitch. *Nat Chem Biol.* **7**: 748-755.

Poiata E, Meyer MM, Ames TD, Breaker RR. 2009. A variant riboswitch aptamer class for *S*-adenosylmethionine common in marine bacteria. *RNA* **15**: 2046-2056.

Price IR, Gaballa A, Ding F, Helmann JD, Ke A. 2015. Mn^{2+}-sensing mechanisms of yybP-ykoY orphan riboswitches. *Mol Cell* **57**: 1110–1123.

Regulski EE, Moy RH, Weinberg Z, Barrick JE, Yao Z, Ruzzo WL, Breaker RR. 2008. A widespread riboswitch candidate that controls bacterial genes involved in molybdenum cofactor and tungsten cofactor metabolism. *Mol Microbiol* **68**:918–932.

Reiss CW, Xiong Y, Strobel SA. 2017. Structural basis for ligand binding to the guanidine-I riboswitch. *Structure* **25**: 195−202.

Ren A, Wang XC, Kellenberger CA, Rajashankar KR, Jones RA, Hammond MC, Patel DJ. 2015. Structural basis for molecular discrimination by a 3',3'-cGAMP sensing riboswitch. *Cell Rep*. **11**: 1-12.

Robbins WJ. 1941. The pyridine analog of thiamine and the growth of fungi. *Proc Natl Acad Sci USA* **27**: 419–422.

Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. 2003. Comparative genomics of the vitamin B_{12} metabolism and regulation in prokaryotes. *J. Biol. Chem.* **278**: 41148–41159.

Roth A, Winkler WC, Regulski EE, Lee BW, Lim J, Jona I, Barrick JE, Ritwik A, Kim JN, Welz R, Iwata-Reuyl D, Breaker RR. 2007. A riboswitch selective for the queuosine precursor preQ1 contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol*. **14**: 308–317.

Seetharaman S, Zivarts M, Sudarsan N, Breaker RR. 2001. Immobilized RNA switches for the analysis of complex chemical and biological mixtures. *Nat Biotechnol*. **19**: 336-341.

Serganov A, Nudler E. 2013. A decade of riboswitches. *Cell* **152**: 17-24.

Serganov A, Yuan Y, Pikovskaya O, Polonskaia A, Malinina L, Phan AT, Hobartner C, Micura R, Breaker RR, Patel DJ. 2004. Structural basis of discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem. Biol*. **11**: 1729–1741.

Shanahan CA, Gaffney BL, Jones RA, Strobel SA. 2011. Differential analogue binding by two classes of c-di-GMP riboswitches. *J Am Chem Soc* **133**: 15578–15592.

Sherlock ME, Breaker RR. 2017. Biochemical validation of a third guanidine riboswitch class in bacteria. *Biochemistry* **56**: 359–363.

Sherlock ME, Malkowski SN, Breaker RR. 2017. Biochemical validation of a second guanidine riboswitch class in bacteria. *Biochemistry* **56**: 352–358.

Sherlock ME, Sadeeshkumar H, Breaker RR. 2019. Variant Bacterial Riboswitches Associated with Nucleotide Hydrolase Genes Sense Nucleoside Diphosphates. *Biochemistry* **58**: 401-410.

Sherlock ME, Sudarsan N, Breaker RR. 2018. Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. *Proc Natl Acad Sci USA* **115**: 6052-6057.

Sherlock ME, Sudarsan N, Stav S, Breaker RR. 2018. Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. *Elife* **7**.

Sherwood AV, Henkin TM. 2016. Riboswitch-mediated gene regulation: Novel RNA architectures dictate gene expression responses. *Annual Review of Microbiology* **70**: 361-374.

Shiota T, Folk JE, Tietze F. 1958. Inhibition of lysine utilization in bacteria by *S*-(beta-aminoethyl) cysteine and its reversal by lysine peptides. *Arch. Biochem. Biophys.* **77**: 372–377.

Smith KD, Lipchock SV, Ames TD, Wang J, Breaker RR, Strobel SA. 2009. Structural basis of ligand binding by a c-di-GMP riboswitch. *Nat Struct Mol Biol*. **16**: 1218-1223.

Smith KD, Strobel SA. 2011. Interactions of the c-di-GMP riboswitch with its second messenger ligand. *Biochem Soc Trans* **39**: 647–651.

Soukup GA, Breaker RR. 2000. Allosteric nucleic acid catalysts. *Curr Opin Struct Biol*. **10**: 318-325.

Spiro S, Guest JR. 1990. FNR and its role in oxygen-related gene expression in *Escherichia coli*. *FEMS Microbiol Rev* **6:** 399–428.

Stormo GD, Ji Y. 2001. Do mRNAs act as direct sensors of small molecules to control their expression? *Proc. Natl. Acad. Sci. USA* **98**: 9465–9467.

Sudarsan N, Cohen-Chalamish S, Nakamura S, Emilsson GM, Breaker RR. 2005. Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem. Biol.* **12**: 1325–1335.

Sudarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, Breaker RR. 2006. Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314**: 300–304.

Sudarsan N, Lee ER, Weinberg Z, Moy RH, Kim JN, Link KH, Breaker RR. 2008. Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **321**: 411–413.

Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. 2003. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev* **17**: 2688–2697.

Swinney DC, Anthony J. 2011. How were new medicines discovered? *Nat. Rev. Drug Discov* **10**: 507–519.

Talbot GH, Bradley J, Edwards JE, Gilbert D, Scheld M, Bartlett JG. 2006. Bad bugs need drugs: an update on the development pipeline from the Antimicrobial Availability Task Force of the Infectious Diseases Society of America. *Clin Infect Dis* **42**: 657–668.

Vicens Q, Mondragón E, Reyes FE, Coish P, Aristoff P, Berman J, Kaur H, Kells KW, Wickens P, Wilson J, Gadwood RC, Schostarez HJ, Suto RK, Blount KF, Batey RT. 2018. Structure-Activity Relationship of Flavin Analogues That Target the Flavin Mononucleotide Riboswitch. *ACS Chem Biol* **13**: 2908-2919.

Vitreschak AG, Lyubetskaya EV, Shirshin MA, Gelfand MS, Lyubetsky VA. 2004. Attenuation regulation of amino acid biosynthetic operons in proteobacteria: comparative genomics analysis. *FEMS Microbiol Lett*. **234**: 357-370.

Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS. 2003. Regulation of the vitamin B_{12} metabolism and transport in bacteria by a conserved RNA structural element. *RNA* **9**: 1084–1097.

Vold B, Szulmajster J, Carbone A. 1975. Regulation of dihydrodipicolinate synthase and aspartate kinase in *Bacillus subtilis*. *J Bacteriol* **121:** 970–974.

Wang H, Mann PA, Xiao L, et al. 2017. Dual-Targeting Small-Molecule Inhibitors of the Staphylococcus aureus FMN Riboswitch Disrupt Riboflavin Homeostasis in an Infectious Setting. *Cell Chem Biol* **24**: 576-588.

Wang JX, Lee ER, Morales DR, Lim J, Breaker RR. 2008. Riboswitches that sense S-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Mol Cell.* **29**:691-702.

Weinberg Z, Barrick JE, Yao Q, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, Sudarson N, Neph S, Tompa M, Ruzzo WL, Breaker RR. 2007. Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* **35**: 4809–4819.

Weinberg Z, Lünse CE, Corbino KA, Ames TD, Nelson JW, Roth A, Perkins KR, Sherlock ME, Breaker RR. 2017a. Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. *Nucleic Acids Res*. **45**: 10811–10823.

Weinberg Z, Nelson JW, Lünse CE, Sherlock ME, Breaker RR. 2017b. Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. *Proc Natl Acad Sci USA* **114**: E2077-E2085.

Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR. 2010. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol* **11**: R31.

Winkler WC, Cohen-Chalamish S, Breaker RR. 2002a. An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci* **99**: 15908–15913.

Winkler WC, Nahvi A, Breaker RR. 2002b. Thiamine derivatives that bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**: 952–956.

Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. 2004. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**: 281-286.

Winkler WC, Navi A, Sudarsan N, Barrick JE, Breaker RR. 2003. An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol* **10**: 701–707.

Wolz C, Geiger T, Goerke C. 2010. The synthesis and function of the alarmone (p) ppGpp in firmicutes. *Int J Med Microbiol* **300**: 142–147.

Woolley DW, White AGC. 1943. Selective reversible inhibition of microbial growth with pyrithiamine. *J Exp Med* **78**: 489–497.

Yao Z, Weinberg Z, Ruzzo WL. 2006. CMfmder—a covariance model based RNA motif finding algorithm. *Bioinformatics* **22**: 445-452.

Zheng W, Thorn, N, McKew JC. 2013. Phenotypic screens as a renewed approach for drug discovery. *Drug Discov Today* **18**: 1067–1073.

Chapter Two

Rare variants of the FMN riboswitch class in *Clostridium difficile* and other bacteria exhibit altered ligand specificity

Largely adapted from the following publication:

Atilho RM, Perkins KR, Breaker RR. 2019. Rare variants of the FMN riboswitch class in *Clostridium difficile* and other bacteria exhibit altered ligand specificity. *RNA* **25:** 23-34.

Summary

Many bacteria use FMN (flavin mononucleotide) riboswitches to control the expression of genes responsible for the biosynthesis and transport of this enzyme cofactor or its precursor, riboflavin. Rare variants of FMN riboswitches found in strains of *Clostridium difficile* and some other bacteria typically control the expression of proteins annotated as transporters, including multidrug efflux pumps. These RNAs no longer recognize FMN, and differ from the original riboswitch consensus sequence at nucleotide positions normally involved in binding of the ribityl and phosphate moieties of the cofactor. In this chapter, my colleagues and I demonstrate that one of the two rare variant subtypes of the FMN riboswitch class bind the FMN precursor riboflavin and the FMN degradation products lumiflavin and lumichrome. Although the biologically relevant ligand sensed by these variant FMN riboswitches remains uncertain, these findings suggest that many strains of *C. difficile* might use rare riboswitches to sense flavin degradation products and activate transporters for their detoxification.

Introduction

Riboswitches are noncoding RNAs (ncRNAs) typically found in the 5′ untranslated regions of bacterial mRNAs, where they regulate gene expression in response to the binding of specific small molecule or ion ligands (Henkin, 2008; Breaker, 2012; Serganov and Nudler, 2013). To date, more than 40 riboswitch classes have been discovered (McCown et al., 2017), many of which control fundamental metabolic processes such as the biosynthesis and transport of amino acids, enzyme cofactors, and nucleotides. In addition, several riboswitch classes have been recently discovered that respond to ligands whose roles in biology were poorly known until recently. These include riboswitches for fluoride (Baker et al., 2012; Ren et al., 2012), guanidine (Nelson et al., 2017; Sherlock et al., 2017a,b), c-di-AMP (Nelson et al., 2013) and c-AMP-GMP (Nelson et al., 2015a; Kellenberger et al., 2015). Thus, the continued identification of novel riboswitch classes provides opportunities to discover hidden or otherwise underappreciated aspects of biology.

Current computational search methods based on comparative sequence analysis (Barrick et al., 2004; Weinberg et al., 2007, 2010, 2017a) have been used to identify numerous widespread riboswitch classes. Unfortunately, the vast majority of the thousands of additional riboswitch classes that likely remain undiscovered (McCown et al., 2017) are probably exceedingly rare, and therefore will be difficult to uncover by using current bioinformatic search techniques. Many of the rare riboswitch classes remaining to be discovered could represent ancient ncRNAs on the brink of evolutionary extinction or they might be niche-specific riboswitches that have more recently emerged within organisms. For example, a newly found SAM-sensing riboswitch, called SAM-VI,

was identified exclusively in *Bifidobacterium* species (Mirihana Arachchilage et al., 2018). Discoveries such as this support the hypothesis that numerous additional rare riboswitch classes likely exist, and that some of these might have emerged through evolution very recently. If the rarest riboswitch classes are largely distinct in their conserved sequence and architectural features, computations approaches that rely on comparative sequence and structure analysis algorithms simply have too few representatives to evaluate.

However, some exceedingly rare riboswitch classes might remain hidden due to their mistaken classification as members of known riboswitch classes. These 'variant' riboswitches might exhibit RNA sequence and structural features that are shared with previously-validated riboswitch classes, but carry key mutations that have altered their ligand specificities (Weinberg et al., 2017). Examples of rare riboswitch classes include variants of guanine riboswitches that sense 2′ -deoxyguanosine (2′-dG). To date, the 2′-dG-I riboswitch class (McCown et al., 2017) has only been identified in *Mesoplasma florum* (Kim et al., 2007). Similarly, the recently discovered 2′-dG-II riboswitch class has only been found within metagenomic sequence data sets (Weinberg et al., 2017). Additional examples include riboswitches for the bacterial alarmone ppGpp (Sherlock et al., 2018a) and the purine biosynthetic precursor phosphoribosyl pyrophosphate (PRPP) (Sherlock et al., 2018b), which were both identified as rare variants of the far more abundant guanidine-I riboswitch class (Nelson et al., 2017).

Riboswitches that bind to flavin mononucleotide (FMN) represent the 7th most abundant riboswitch class that has been experimentally validated to date (McCown et al., 2017). Members of this common riboswitch class usually control the expression of genes

required for the biosynthesis and transport of the coenzyme precursor riboflavin (Gelfand et al., 1999; Mironov et al., 2002; Winkler et al., 2002). FMN and flavin adenine dinucleotide (FAD) are ordinarily used as redox cofactors for flavoenzymes involved in various aspects of cellular metabolism (Fischer and Bacher, 2005). However, due to the high photoreactivity of these compounds, they can degrade into various toxic byproducts such as lumiflavin and lumichrome (Fall and Petering, 1956; West and Owen, 1973; Choe et al., 2005). Some bacterial and archaeal cells protect themselves from these phototoxic products with riboflavin-binding proteins called dodecins that act to both store riboflavin and prevent its undesirable breakdown (Grininger et al., 2006, 2009). Presumably some bacterial species also have detection and remediation systems in the event that substantial amounts of FMN breakdown products accumulate.

In this chapter, my colleagues and I present our analysis of two classes of variant FMN riboswitches whose initial members were previously discovered in various strains of *Clostridium difficile* (Blount et al., 2012, 2013; Weinberg et al., 2017b) and originally called *CD3299* RNA. Furthermore, bioinformatic analysis has yielded additional representatives of these variant FMN motif RNAs in strains of *Atopobium* and in bacterial metagenomic DNA sequences. We describe two highly similar variant subtypes, and demonstrate that all representatives tested completely reject FMN. Intriguingly, I demonstrate that subtype 2 representatives strongly bind riboflavin and certain light-mediated FMN degradation products. We hypothesize that some variant FMN RNAs might have evolved to recognize flavin derivatives to activate transporters involved in their disposal. These findings reinforce the hypothesis that many riboswitch classes remaining to be discovered will be exceptionally rare and challenging to solve.

Results

Variant RNAs have key characteristics that differ from typical FMN riboswitches

Examples of FMN riboswitch variants were previously reported (Blount et al., 2012, 2013) to exist only in several strains of *C. difficile*. Additional examples were subsequently uncovered in a bioinformatics study that purposefully examined known riboswitch classes for the existence of variant riboswitch aptamers that have possibly altered their ligand binding specificities (Weinberg et al., 2017b). Given the ongoing rapid expansion of bacterial genomic sequence datasets, Kevin Perkins searched (see Materials and Methods) for additional representatives of the FMN variants and identified a total of 57 unique-sequence examples present in certain *Clostridium* and *Atopobium* species, as well as in bacterial metagenomic DNA sequence data (RefSeq version 76 and additional environmental DNA sequence databases).

These findings allowed us to create revised consensus sequence models for the variants, and to increase the list of genes known to be associated with these RNAs. Specifically, the variants have been organized into two groups, called subtype 1 (25 examples) (**Figure 2-1A**) and subtype 2 (32 examples) (**Figure 2-1B**). In addition, Kevin Perkins updated the collection of FMN riboswitch representatives from the same genomic and metagenomic sequence databases to establish a current consensus model for this well-established riboswitch class (**Figure 2-1C**). The variants exhibit many similarities, both with regards to their conserved nucleotide sequences and their major secondary structure elements, to the FMN riboswitch consensus model. Due to these abundant similarities, it is not surprising that the rare variants previously had been grouped with the much larger collection of FMN riboswitch representatives (11,603 examples), including

in the Rfam Database (Nawrocki et al., 2015). This lumping of multiple riboswitch classes into a single collection can occur when variants do not substantially differ from the established aptamer consensus, or when too few representatives exist for their discovery by careful comparison of nucleotide sequences.

Despite these commonalities, the differences between authentic FMN riboswitches and the two variant subtypes are striking. The distinct features of the variants can be best appreciated by comparison with the gene associations, conserved sequences, and structural features of FMN-sensing riboswitches. FMN riboswitches usually reside upstream of genes involved in riboflavin biosynthesis (Gelfand et al., 1999; Winkler et al., 2002; Mironov et al., 2002) (**Figure 2-1C**, bottom). Genes encoding COG3601, COG0697, COG3201, and TauC family proteins are also commonly associated with FMN riboswitches. These proteins have been characterized as riboflavin transporters in various organisms (Vogl et al., 2007; García Angulo et al., 2013; Gutiérrez-Preciado et al., 2015).

In stark contrast to FMN riboswitch regulation of genes involved in the biosynthesis and import of riboflavin, subtype 1 variant FMN riboswitches are never associated with FMN biosynthetic genes, but are currently exclusively observed in association with genes for members of the COG3601 transporter family (**Figure 2-1A**, bottom). Curiously, genes annotated as *COG3601* are also commonly regulated by true FMN riboswitches and, as noted above, in those instances are predicted to be riboflavin transporters (Gutiérrez-Preciado et al., 2015). However, there are many examples of *COG3601* genes that are linked to other metabolic processes, and the specificities of those transporters are proposed to be different (Gutiérrez-Preciado et al., 2015). Thus, it seems possible that the

COG3601 genes whose expression is regulated by subtype 1 RNAs might have a distinct substrate compared to those regulated by true FMN riboswitches.

Likewise, the subtype 2 variant FMN riboswitches are exclusively found upstream of genes encoding putative multidrug efflux transporters (**Figure 2-1B**, bottom). In various strains of *C. difficile* subtype 2 variant FMN riboswitches are commonly associated with genes coding for proteins annotated as putative EmrB and NorM multidrug transporters from the major facilitator superfamily (MFS) and multidrug and toxic compound extrusion (MATE) superfamily, respectively (**Figure 2-1B**, right). Certain members of the MATE superfamily, such as the *Vibrio parahaemolyticus* protein NorM, have been demonstrated to transport a wide variety of cationic dyes, quaternary ammonium compounds, and quinolones (Morita et al., 1998). Moreover, transporters associated with variant FMN RNAs have been previously shown to contain MFS 1 domains (Gutiérrez-Preciado et al., 2015). Proteins that contain these domains exhibit homology to the *Escherichia coli* protein EmrB. Members of this protein class are part of the EmrAB-TolC tripartite system, which has been shown to confer resistance to hydrophobic toxins (Lomovskaya and Lewis, 1992; Lewis, 2000).

Not a single example of a *norM* or *emrB* gene is associated with the many thousands of FMN riboswitches that have been identified by using bioinformatics search methods. Again, the patterns of gene associations between true FMN riboswitches and these rarer variants strongly indicate that the ligand for the variant riboswitches has been changed from FMN to another, but perhaps structurally related, compound. Presumably, the natural substrate for these specific transporters will match the ligand for the variant riboswitches. Unfortunately, the precise ligands transported by the proteins whose genes

are regulated by the variant riboswitches are not yet known. Indeed, as described in greater detail later in this report, I used ligand-binding assays to demonstrate that these variants do not recognize the cofactor FMN (**Figure 2-4**). These findings are consistent with our hypothesis that the variants do not function as FMN-responsive riboswitches, but carry nucleotide differences that have altered their ligand specificities.

Most importantly, the nucleotide sequence differences between true FMN riboswitches and the rare variants almost exclusively reside in otherwise highly conserved nucleotides that are known to form direct contacts with the FMN ligand (**Figure 2-2**). Previously, the FMN riboswitch aptamer from the *Bacillus subtilis ribD* mRNA was found to exhibit a dissociation constant (K_D) for FMN of 5 nM (Winkler et al., 2002). X-ray crystallography was subsequently performed with another member of the FMN riboswitch class from *Fusobacterium nucleatum* (Serganov et al., 2009). FMN is recognized by the RNA predominantly through interactions with the isoalloxazine ring and through contacts with the phosphate on the ribityl moiety (**Figure 2-2A**). The isoalloxazine ring system intercalates between the bases of two adenosine nucleotides (A48 and A85) (**Figure 2-1C**) of the aptamer, and also forms two hydrogen bonds with another highly conserved adenosine (A99) (**Figure 2-2A**). The ribityl and phosphate moieties are selectively recognized by the formation of numerous hydrogen bonds with several conserved guanosine nucleotides distributed throughout the riboswitch aptamer (G10, G11, G32, G62, G84) (**Figure 2-2A**). The absence of the phosphate moiety, such as with riboflavin, has been found to reduce ligand affinity by almost three orders of magnitude (Winkler et al., 2002), which demonstrates the importance of this set of highly-conserved guanosine nucleotides.

The two variant FMN RNA subtypes (**Figure 2-1A, B**) have many sequence and structural features resembling those of the FMN riboswitch consensus (**Figure 2-1C**), but importantly they deviate from the consensus sequence at certain nucleotide positions that are directly involved in recognizing FMN. These changes to the ligand binding site typically convert the highly conserved (>97%) guanosine nucleotides that are known to contact the phosphate and ribityl moieties of FMN (**Figure 2-2A**) to adenosine nucleotides (**Figure 2-2B**). Presumably, the variant RNAs no longer have the ability to create productive binding interactions with the phosphate moiety of FMN, which explains why the variant RNAs examined for FMN ligand binding completely reject this enzyme cofactor (see below). However, the conservation of nucleotide identities at positions equivalent to 48, 85 and 99, which contact the flavin moiety, suggest that this part of the natural ligand for subtypes 1 and 2 RNAs might remain unchanged.

There are some notable sequence and structural features that also differ between the two subtypes. Most obvious is the presence of an extended P3 stem (called P3a and P3b) in the consensus model for subtype 1 RNAs (**Figure 2-1B**) that is lacking in subtype 2 RNAs (**Figure 2-1B**). However, nearly half of all true FMN riboswitches carry an extended P3 stem, and so it is not certain that this distinction affects the ligand specificity of subtype 1 RNAs. However, there are modest differences in the key nucleotides that form the original FMN binding site between subtype 1 and 2 RNAs (**Figure 2-2B**). Given the differences in gene associations between these two RNAs, it seems possible that these nucleotide changes also result in different ligand specificities.

Subtype 2 variant FMN RNAs recognize photolytic products of FMN and FAD

Frequently, the ligand that is sensed by a riboswitch class is identified by examining the genes that reside immediately downstream. Unfortunately, as noted above, the selectivities of the transporter proteins encoded by the genes downstream of the variant FMN RNAs remain unknown, and thus provide limited clues regarding the identity of the cognate ligand. Because many *C. difficile* strains already carry a gene for a putative riboflavin importer whose expression is controlled by an FMN riboswitch, we hypothesized that some bacteria might use variant FMN RNAs to control the expression of efflux pumps to alleviate the deleterious build-up of toxic FMN breakdown products (Fall and Petering, 1956; West and Owen, 1973). This would be analogous to the observed use of guanidine riboswitches for controlling the production of multidrug efflux pumps (Nelson et al., 2017, Kermani et al., 2018). Both FMN and riboflavin are highly photoreactive and have been shown to generate various light-promoted breakdown products such as lumiflavin, lumichrome and carboxymethylflavin (Choe et al., 2005). Small riboflavin-binding proteins called dodecins are used by archaeal and bacterial cells to prevent flavin derivatives from damaging key molecules in the cell by stabilizing riboflavin against photochemical breakdown or by sequestering toxic byproducts, such as lumichrome (Grininger et al. 2006, 2009).

To assess the ability of variant FMN RNAs to directly bind photochemical breakdown products of FMN, my colleague Kevin Perkins and I performed in-line probing assays with a 134 nucleotide RNA, named 134 *norM*, which is derived from the 5′-untranslated region of the *norM* gene of *Clostridium sp. ASF356* (**Figure 2-3A**). In-line probing reveals ligand-mediated shape changes in RNA structures by monitoring the

products of spontaneous RNA phosphodiester cleavage (Soukup and Breaker, 1999). Kevin Perkins observed that both UV-treated FMN and UV-treated FAD cause decreased levels of spontaneous RNA strand scission at various locations of the 134 *norM* RNA chain, suggesting that breakdown products of FMN are bound by this putative aptamer (**Figure 2-4A**).

Indeed, reproducible changes in the banding pattern produced by in-line probing incubations are also exhibited by the 134 *norM* RNA in the presence of riboflavin, lumiflavin, and lumichrome (**Figure 2-4B**), when each was individually tested at 100 μM (**Figure 2-3B**). I also observed similar ligand-induced structural modulation of subtype 2 RNA constructs from other organisms (**Figure 2-5, 2-6**). These changes to the RNA structure as inferred from differences in the banding patterns produced by the in-line probing assays are occurring at nucleotide positions that are at or immediately adjacent to those corresponding to the binding site nucleotides for FMN riboswitches (**Figure 2-1, Figure 2-2A**). In contrast, I determined that FMN is rejected by the 134 *norM* construct and the other subtype 2 RNAs examined, presumably because most of the original nucleotides of the ribityl-phosphate binding site are different in subtype 2 RNAs (**Figure 2-2B**). I demonstrated that the 134 *norM* RNA binds lumichrome and lumiflavin more tightly than it does riboflavin (**Figure 2-3C**, **Figure 2-7**), suggesting that the chemical structure of the ligand that replaces the ribityl and phosphate moieties of FMN are important for binding. This is consistent with our prediction that the extensive changes in the identities of nucleotides forming the original ribityl-phosphate binding site that occur with subtypes 1 and 2 RNAs would affect recognition of that portion of the ligand, but would not necessarily affect binding of the flavin moiety.

The in-line probing data for 134 *norM* also indicate that the various ligands bind with a 1:1 interaction (**Figure 2-3C**, **Figure 2-7**). Similar results are exhibited by another subtype 2 RNA construct called 136 *emrB* RNA derived from *C. difficile CD196* (**Figure 2-8**). My findings are consistent with the data observed previously for FMN riboswitches (Winkler et al., 2002; Serganov et al., 2009), suggesting that the nucleotide differences in the binding site of FMN riboswitches compared to subtype 2 RNAs cause a change in ligand specificity, but not stoichiometry.

Unlike subtype 2 RNA constructs, my colleagues and I have not yet observed ligand binding by subtype 1 RNA constructs. I used in-line probing assays to examine the function of a 186-nucleotide construct associated with the *COG3601* gene from the bacterium *Atopobium sp. BS2*, called 186 *COG3601* (**Figure 2-9**). The construct fails to respond to the addition of 100 μM FMN, which suggests that these RNAs have also altered their ligand specificity. However, both UV-treated samples and known photochemical breakdown products of FMN fail to induce any substantial structural modulation. Also tested was the FMN analog 10-(3-(4-fluorophenyl)propyl)-7,8-dimethylbenzo[g]pteridine-2,4(3H,10H)-dione (called 10FDPD) (Blount et al., 2012, 2013), which contains an aryl-alkyl moiety in place of the ribityl phosphate moiety (**Figure 2-4B**). Compounds like 10FDPD were developed as FMN analogs that function as antibacterial agents that target FMN riboswitches (Blount et al., 2015), and 10FDPD modulates the structure of a subtype 2 RNA construct (**Figure 2-4A**). Again, unlike subtype 2 RNAs, the subtype 1 RNA construct 186 *COG3601* fails to bind this FMN analog. It is not known whether subtype 1 RNAs recognize a natural ligand that is

different from that of the other variants, or whether the construct tested in vitro fails to respond as they do within their native biological context.

Subtype 2 RNAs from *Clostridium* species are genetic "ON" riboswitches

The *B. subtilis ribD* FMN riboswitch controls the expression of the entire riboflavin biosynthesis operon through transcription termination regulation (Winkler et al., 2002). When FMN is bound by the riboswitch, an intrinsic transcription terminator stem (Gusarov and Nudler, 1999; Yarnell and Roberts, 1999) forms that halts extension of the nascent mRNA transcript. In the absence of FMN, an anti-terminator structure can form to prevent formation of the terminator. Thus, the competing terminator and anti-terminator structures constitute the 'expression platform' (Breaker, 2012) of the riboswitch. A second FMN riboswitch representative also found in *B. subtilis* controls expression of the *ribU* gene (previously called *ypaA*), which encodes a riboflavin transporter of the COG3601 protein family (Vogl et al., 2007). This riboswitch is predicted to regulate translation initiation by preventing ribosome access to the ribosome binding site (Shine-Dalgarno sequence) upon binding to FMN (Winkler et al., 2002). These examples provided by FMN riboswitches from the same organism showcase two of the most common expression platform mechanisms by which riboswitches control gene expression in response to the accumulation of their target metabolites (Breaker 2012).

Regardless of the mechanism used, riboswitches almost always incorporate an expression platform downstream of the aptamer to regulate the level of gene expression upon ligand binding. Therefore, to assess how variant FMN riboswitches control gene

expression, I manually inspected sequences downstream of several variant FMN RNA aptamers for evidence of expression platform structures. Subtype 2 RNAs, which associate with genes annotated as multidrug transporters, typically precede a putative intrinsic terminator stem. These subtype 2 aptamers therefore most probably regulate gene expression via transcription termination, and should activate the expression of genes for efflux pumps if they indeed recognize toxic FMN breakdown products.

To experimentally assess the mechanism of variant 2 RNAs most similar to those found in *C. difficile*, Kevin Perkins conducted single-round transcription termination assays (Landick et al., 1996) using a DNA template derived from the region upstream of the *norM* gene from *C. sp. ASF356.* The RNA derived from this *norM* construct has the potential to form a strong antiterminator stem within the expression platform (orange shading, **Figure 2-10A**). Folding of the aptamer structure in the presence of ligand should permit formation of this antiterminator stem and thereby promote transcription of the full-length mRNA. In the absence of ligand, alternative folding appears possible because nucleotides in the right shoulder of P2 are complementary to nucleotides required to form the antiterminator. Formation of this putative 'anti-antiterminator' stem in the absence of ligand should halt transcription by allowing the formation of the intrinsic terminator stem.

The amounts of terminated and full-length RNA transcripts generated in the presence of FMN and several analogs were measured with the *C. sp. ASF356 norM* DNA template, and with a template carrying an FMN riboswitch derived from the *ribD* gene from *C. difficile* 630 (**Figure 2-10B**). Kevin Perkins observed modest changes in the fraction of full-length transcripts when riboflavin, lumiflavin, and particularly lumichrome are added

to single-round in vitro transcription reactions (**Figure 2-10B**, left). Lumichrome also has the best observed K_D value I determined by in-line probing assays for these variant RNAs (**Figure 2-3C**). In contrast, no changes in the fraction of elongated transcripts are seen for the *ribD* FMN riboswitch when these same compounds are introduced (**Figure 2-10B**, right). However, as expected, FMN does alter the relative transcription yields for this natural FMN riboswitch, whereas FMN has no effect on the distribution of *norM* RNA transcripts.

I examined another subtype 2 riboswitch representative, this one derived from the *emrB* gene of *C. difficile* 630, for in vivo gene control activity by fusing the construct or several mutant versions of this construct to a *lacZ* β-galactosidase reporter gene (**Figure 2-11A**). When *B. subtilis* cells carrying the wild-type (WT) *emrB* variant construct are cultured in rich (lysogeny broth, LB) medium liquid, reporter gene expression is negligible (**Figure 2-11B**). In contrast, when mutations (construct M1) are introduced to disrupt the predicted intrinsic terminator stem, robust reporter gene expression occurs. Furthermore, disruption of key conserved RNA sequence or structural features that form the aptamer, such as in mutant reporter constructs M2, M3 and M4, fails to change reporter gene expression. These results suggest that, under these culture conditions, the WT riboswitch is in its "OFF" state and that ligand is not present at concentrations sufficient to be bound to the aptamer. Reporter construct M5 and M6 do exhibit modest gene expression increases relative to the WT construct, which is not entirely consistent with our hypothesis that subtype 2 RNAs function as "ON" switches. However, we speculate that these mutations might cause folding problems that interfere somewhat with

terminator stem formation, rather than serve as evidence for "OFF" switch function by this riboswitch representative.

Similarly, the *C. difficile 630* subtype 2 riboswitch reporter construct in *B. subtilis* cells grown on LB medium agar plates exhibits no apparent signs of gene expression (**Figure 2-11C**). Agar-diffusion assays revealed that riboflavin, lumichrome, and alloxazine also do not trigger expression of the reporter construct in *B. subtilis*. However, all three compounds lack zones of cell growth inhibition, meaning that they do not accumulate to toxic levels inside *B. subtilis* cells. Thus, we cannot be certain that concentrations necessary to trigger riboswitch function can be achieved by supplying these molecules in the culture medium. Although reporter gene expression (blue color) is seen with the addition of lumiflavin to a filter disk, reporter constructs M2 through M4 failed to disrupt reporter gene expression (data not shown). This result is most likely caused by non-specific activation of our reporter gene, as lumiflavin was also seen to modestly activate reporter gene expression when fused to FMN riboswitches (data not shown). This false positive signal might occur when toxic compounds affect overall mRNA transcription or translation levels.

My colleagues and I did not conduct parallel studies on a representative of subtype 1 variant riboswitches, in part due to the fact that we have not observed binding for any of the candidate ligands tested in vitro. In contrast to subtype 2 RNAs, Kevin Perkins determined that all subtype 1 riboswitches located upstream of genes encoding COG3601 family proteins in *Atopobium* species are predicted to translationally control gene expression via sequestration of ribosome binding site. Formation of the P1 stem of the variant aptamer in the presence of ligand would preclude access of the ribosome binding

site and repress gene expression. These RNAs thus appear to repress the expression of their associated transporters in response to binding by their natural ligand, which is opposite of that predicted for subtype 2 variants. Either subtype 1 RNAs have yet another ligand specificity, or cells that carry these RNAs benefit by repressing the expression of their associated transporters whereas subtype 2 variants activate the expression of their associated transporters.

Variant FMN riboswitches hint at an unknown flavin-like detoxification system

Toxic compounds often build up from the spontaneous chemical breakdown of coenzymes and cofactors, or from enzymatic reactions that yield unwanted side products (Linster et al., 2013). Flavins, such as FMN and FAD, which are commonly utilized by flavoproteins to catalyze a wide range of redox reactions, are no exception from this problem particularly due to their high photoreactivity (Fischer and Bacher, 2005). Various types of damage control systems are utilized by bacteria and other species to prevent the harmful build-up of toxic agents. For example, many bacteria utilize riboswitches to sense toxic ligands, such as *S*-adenosylhomocysteine (Wang et al., 2008), fluoride (Baker et al., 2012), or guanidine (Nelson et al., 2017), and to activate genes important for toxicity resistance.

It also seems possible that subtype 2 FMN RNAs activate the production of efflux proteins to alleviate the harmful buildup of their cognate ligand. Consistent with this hypothesis is the observation that the subtype 2 *norM* riboswitch includes an expression platform with terminator, antiterminator, and possible anti-antiterminator base-pairing potential (**Figure 2-10A**). Indeed, lumichrome, which was found to bind most tightly to

representative subtype 2 RNAs (**Figure 2-3C**), promotes elongation of the *norM* riboswitch construct in single-round transcription termination assays (**Figure 2-10B**, left).

Unfortunately, due to the limited genetic contexts and limited analysis of ligand analogs, we cannot yet conclude that any of these molecules are the natural ligand for the subtype 2 riboswitch class. As noted above, many of the genes predicted to be controlled by subtype 2 RNAs are annotated as putative multidrug transporters. However, one of these genes from *C. difficile* has been recently proposed to code for a riboflavin import protein (Gutiérrez-Preciado et al., 2015). It is important to note that multidrug resistance transporters have also been observed to function inefficiently in the opposite direction by bringing molecules into the cell (Robinson et al., 2017), and many of the *C. difficile* strains that carry variant FMN RNAs often already carry another gene regulated by an FMN riboswitch that codes for a riboflavin importer. Therefore, the natural function of the transporters associated with the variant riboswitches might involve the efflux of toxic flavin compounds, rather than the import of riboflavin.

Consistent with this hypothesis is Kevin Perkins observation that ligand binding activates production of the full-length RNA transcript, which is opposite of what cells would need if riboflavin were the ligand, and the associated transporter were riboflavin importers. Transporters for fluoride (Baker et al., 2012; Stockbridge et al., 2013) and guanidine (Nelson et al., 2017; Kermani et al., 2018) are two examples of proteins whose annotations have been changed in recent years as orphan riboswitch classes have been solved. Thus, the identification of the true natural ligand for the FMN riboswitch variants described herein would also help establish the substrates targeted by the transporters.

Due to the lack of compound availability, we were unable to generate sufficient data to formulate a structural model for the cognate ligand for variant FMN RNAs using structure-activity relationship (SAR) assays. Many of the predicted photochemical breakdown products of FMN and FAD contain a variation of the ribityl tail (**Figure 2-4C**). Although subtype 2 variant riboswitch RNAs bind lumichrome with the tightest binding affinity measured in this study (**Figure 2-3C**), drug-like FMN analogs similar to 10FDPD have been previously identified to bind FMN riboswitches just as tightly as the cognate ligand (Blount et al., 2015). Thus, we cannot be certain that tight-binding ligand analogs are perfect guides to the chemical structure of the true ligand. Also, lumichrome reduces the amount of terminated product during in vitro transcription assays, but fails to inhibit *B. subtilis* cell growth or trigger reporter gene expression. Perhaps the cognate ligand for variant FMN RNAs is a distinct compound endogenous to *C. difficile*. Or, perhaps the *B. subtilis* surrogate organism used in our study to evaluate reporter gene regulation already carries an efficient detoxification pathway to deal with lumichrome toxicity, which prevents our in vivo gene regulation assays from revealing gene regulation by this compound.

Discussion

In the current study, my colleagues and I provide data on variant riboswitches that carry mutations in the ligand-binding aptamer domain of the FMN class. Most notably, these variant RNAs carry mutations of nucleotide positions that directly recognize the ribityl and phosphate moieties, and so the mechanism for the loss of FMN recognition is readily apparent. Kevin Perkins and I demonstrate that degradation products of FMN, such as riboflavin, lumiflavin, and lumichrome, are recognized by certain variant FMN RNAs of the subtype 2 group (**Figure 2-3**). However, due to obscure gene associations, it is not immediately evident what cognate ligand this variant riboswitch class senses. Previous work has determined that it is possible to design synthetic analogs that trick riboswitch classes into altering gene expression (Kim et al., 2009; Howe et al., 2015; Blount et al., 2015). Thus, the isoalloxazine derivatives might only be close mimics of the cognate ligand, and the true cognate ligand might await discovery.

The expression platform for subtype 2 variant RNAs is predicted to activate gene expression in the presence of ligand (**Figure 2-10**). This logic is similar to other riboswitch classes that sense toxic ligands and activate the expression of transporter proteins (Baker et al., 2012; Nelson et al., 2017; Sherlock et al., 2017a,b). Consistent with this hypothesis is the fact that these variant RNAs are commonly found upstream of genes annotated as multidrug transporters. These transporters have similarity to members of the EmrB/QacA family of MFS exporters whose typical substrates are polyaromatic cations. In other Gram-positive bacteria, QacA transporters are regulated by the QacR protein, which functions as a transcriptional repressor of the *qacA* gene (Grkovic et al., 1998). When substrates for QacA accumulate in cells, they bind to QacR and prevent it

from binding DNA, thereby increasing expression levels of QacA transporters. If the sensory and regulatory functions of subtype 2 variant FMN RNAs are analogous to those of QacR, then these RNAs most likely function to turn on gene expression in response to toxic flavin metabolites.

Many flavin compounds, such as riboflavin, are highly photoreactive, and can become harmful for cells when they accumulate or degrade into various phototoxic byproducts (Choe et al., 2005). However, in the riboflavin overproducing strain, *Ashbya gossypii*, riboflavin overproduction during sporulation has been shown to protect spores from ultraviolet (UV) light (Stahmann et al., 2001). Therefore, anaerobic spore-forming bacteria, such as *C. difficile*, might utilize variant FMN riboswitches to activate detoxification pathways necessary to efflux toxic flavin compounds that accumulated as a spore. Since *C. difficile* is a major cause of nosocomial infections, understanding the mechanism by which *C. difficile* spores are highly resistant to environmental stresses such as UV could aid in the development of new therapeutic or decontamination methods.

Material and Methods

Chemicals, DNA oligonucleotides, and bacterial strains. Chemical compounds were purchased from Sigma-Aldrich with the exception of 10FDPD, which was acquired from the former assets of BioRelix. [γ-^{32}P]-ATP and [α-^{32}P]-ATP were purchased from PerkinElmer. DNA oligonucleotides (**Table 2-1**) were purchased from Sigma-Aldrich or Integrated DNA Technologies. All enzymes were purchased from New England BioLabs unless otherwise specified. *B. subtilis* strain 168 (BGSC 1A747) was obtained from the Bacillus Genetic Stock Center (BGSC) at The Ohio State University.

Bioinformatic analyses. Additional homologs of the variant FMN motif RNAs were identified using Infernal 1.1 (Nawrocki and Eddy, 2013). Searches were conducted with RefSeq (O'Leary et al., 2016) version 76 and various metagenomic databases as previously described (Weinberg et al., 2017b). Sequence and secondary structure consensus model of the variant FMN motif RNAs (57 sequences) were constructed using R2R (Weinberg and Breaker, 2011).

Preparation of RNA oligonucleotides. RNAs were prepared by in vitro transcription using DNA oligonucleotides containing a T7 RNA polymerase promoter sequence upstream of the desired template sequence (**Table 2-1**). Specifically, in vitro transcription reactions were performed using laboratory-prepared bacteriophage T7 RNA polymerase (2 units μL^{-1}) in 80 mM HEPES (pH 7.5 at 23°C), 40 mM DTT, 24 mM $MgCl_2$, 2 mM spermidine, and 2 mM of each NTP. RNA was purified using denaturing (8 M urea) 10% polyacrylamide gel electrophoresis (PAGE). RNA bands were visualized by UV

shadowing, excised, and then eluted from the gel slice by the crush-soak method overnight in 10 mM Tris-HCl (pH 7.5 at 23°C), 200 mM NaCl, and 1 mM EDTA. The RNA was precipitated by adding two volumes of 100% ethanol and incubating at -20°C for 30 min, and then pelleted by centrifugation.

To generate RNAs ^{32}P-radiolabeled at the 5′-terminus, RNA was subjected to dephosphorylation using rAPid Alkaline Phosphatase (Roche Diagnostics), and then radiolabeled with [γ-^{32}P]-ATP at the 5′-terminus using T4 polynucleotide kinase. Radiolabeled RNAs were purified by denaturing 10% PAGE and recovered as described above.

In-line probing of RNAs. In-line probing assays were performed as previously described (Regulski and Breaker, 2008; Soukup and Breaker, 1999). Briefly, 5′ ^{32}P-labeled RNAs were incubated in the presence or absence of different concentrations of desired ligand at 23°C for 42 to 48 h in the presence of 20 mM $MgCl_2$, 100 mM KCl, and 50 mM Tris-HCl (pH 8.3 at 23°C). The spontaneous RNA cleavage products were separated by denaturing 10% PAGE. The gels were dried and imaged with a Typhoon phosphorImager (GE Healthcare). Dissociation constants were established by varying the concentration of ligand added and quantifying the changes in band intensities via ImageQuaNT at nucleotide positions that exhibited ligand-induced structural modulation. Values for band intensities were normalized to a non-modulating band, scaled between 0 and 1, and then plotted as a function of the logarithm of the ligand concentration. Values for apparent dissociation constants were determined using a sigmoidal dose-response equation and GraphPad Prism 7.

In vitro transcription termination assays. The protocol for single-round in vitro transcription was adapted from one previous descriptions (Landick et al., 1996). DNA constructs were designed to include the promoter sequence of the *lysC* gene from *B. subtilis*, the riboswitch aptamer, and the expression platform of the *ribD* gene or *emrB* gene from *C. difficile* to 57 or 42 nucleotides, respectively, following the terminator stem. Approximately 2 pmol of the purified, PCR amplified DNA template was added to a transcription initiation mixture (20 mM HEPES [pH 8.0 at 23ºC], 5 mM magnesium acetate, 4 mM DTT, 1 mM EDTA, 10 μg mL^{-1} bovine serum albumin [BSA], 130 μM ApA dinucleotide, 1% glycerol, 0.04 U μL^{-1} *E. coli* RNA polymerase holoenzyme, 2.5 μM GTP, 2.5 μM UTP, and 1.5 μM ATP). Approximately 8 μCi [α-^{32}P]-ATP was added to the 90 μL transcription reaction and transcription was allowed to proceed at 37ºC for 10 minutes, leading to formation of a stalled polymerase complex at the first cytidine nucleotide of each transcript. The reaction mixture was then distributed in 8 μL aliquots into separate tubes, which contained 1 μL of a 10x solution of the ligand of interest and 1 μL of 10x elongation buffer (20 mM HEPES [pH 8.0 at 23ºC], 5 mM magnesium acetate, 4 mM DTT, 1 mM EDTA, 1 mg mL^{-1} heparin, 75 μM each of ATP, GTP, and CTP, and 25 μM UTP). Products were separated by denaturing 10% PAGE then imaged and quantified using a Typhoon phosphorImager and ImageQuaNT software. The band intensities for both the full-length (FL) and terminated (T) RNA transcripts were measured, and the fraction FL values were calculated by using the equation Fraction FL = (FL intensity)/(FL intensity + T intensity). This simple equation can be used because the differences in specific activities between the FL and T products due to [α-^{32}P]-ATP incorporation are negligible.

Design of reporter gene constructs. The nucleotide region –299 to +1 relative to the *emrB* translation start site of the downstream gene was amplified by PCR from *C. difficile* 630 genomic DNA. Subsequently, the *B. subtilis lysC* promoter was introduced via PCR amplification. This fragment was cloned into the vector pDG1661 as a transcriptional fusion with a *lacZ* reporter gene. The resulting plasmid was transformed and integrated into the *amyE* locus of *B. subtilis* 1A1.

Agar diffusion assays. *B. subtilis* strains carrying riboswitch reporter constructs were grown on LB agar plates containing X-gal (100 μg mL^{-1}) and appropriate antibiotic. Autoclaved 6 mm diameter pap er discs prepared from 0.35 mm thick pure cellulose chromatography paper (Fisher Scientific) were soaked with 10 μL of compound at specific concentrations and transferred to prepared agar plates. The plates were incubated overnight at 37°C prior to analysis.

Liquid-based β-galactosidase assays. Liquid-based β-galactosidase assays were performed as previously described (Nelson et al., 2015b; Vidal-Aroca et al., 2006). Bacterial cell cultures were grown overnight in 3 mL of LB medium with appropriate antibiotics at 37°C with shaking. The next day, cells were diluted 1:10 in LB and 80 μL of the resulting cultures were transferred to Costar black 96-well clear-bottom plates. The resulting samples were incubated for 4 hours at 37°C with shaking. The absorbance at 595 nm was measured using a Tecan Infinite M200 PRO microplate reader. Subsequently, 80 μL of Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1

mM $MgSO_4$) and 40 µL of 1 mg mL^{-1} 4-methylumbelliferyl-β-D-galactopyranoside (4-MUG) were added to each well. The mixture was allowed to incubate at room temperature for 15 min, and 40 µL of 1 M Na_2CO_3 was used to quench the reaction. Excitation and emission values were measured at 360 nm and 460 nm respectively, using a Tecan Infinite M200 PRO microplate reader. Fluorescence units were calculated as previously described (Vidal-Aroca et al., 2006).

References

Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, Breaker RR. 2012. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335:** 233–235.

Barrick JE, Corbino KA, Winkler WC, Nahvi A, Mandal M, Collins J, Lee M, Roth A, Sudarsan N, Jona I, et al. 2004. New riboswitch motifs suggest an expanded scope for riboswitches in bacterial genetic control. *Proc Natl Acad Sci USA* **101:** 6421–6426.

Blount KF. 2013. Methods for treating or inhibiting infection by *Clostridium difficile.* U.S. patent appl. no. 13/576,989.

Blount KF, Coish PDG, Dixon BR, Myung J, Osterman D, Wickens P, Avola S, Baboulas N, Bello A, Berman J, et al. 2012. Flavin derivatives. U.S. patent appl. no. 13/381,809.

Blount KF, Megyola C, Plummer M, Osterman D, O'Connell T, Aristoff P, Quinn C, Chrusciel RA, Poel TJ, Schostarez HJ, Stewart CA, Walker DP, Wuts PG, Breaker RR. 2015. Novel riboswitch-binding flavin analog that protects mice against *Clostridium difficile* infection without inhibiting cecal flora. *Antimicrob Agents Chemother* **59:** 5736–5746.

Breaker RR. 2011. Prospects for riboswitch discovery and analysis. *Mol Cell* **43:** 867–879.

Breaker RR. 2012. Riboswitches and the RNA world. *Cold Spring Harb Perspect Biol* **4:** a003566.

Choe E., Huang R. & Min DB. 2005. Chemical reactions and stability of riboflavin in foods. *J Food Sci* **70:** R28–R36.

Fall HH, Petering HG. 1956. Metabolite Inhibitors. I. 6, 7-Dimethyl-9-formylmethyl-isoalloxazine, 6, 7-Dimethyl-9-(2'-hydroxyethyl)-isoalloxazine and Derivatives. *J Am Chem Soc* **78:** 377–380.

Fischer M, Bacher A. 2005. Biosynthesis of flavocoenzymes. *Nat Prod Rep* **22:** 324–350.

García Angulo VA, Bonomi HR, Posadas DM, Serer MI, Torres AG, Zorreguieta A, Golbaum FA. 2013. Identification and characterization of RibN, a novel family of riboflavin transporters from *Rhizobium leguminosarum* and other proteobacteria. *J Bacteriol* **195:** 4611–4619.

Gelfand MS, Mironov AA, Jomantas J, Kozlov YI, Perumov DA. 1999. A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes. *Trends Genet* **15:** 439–442.

Grininger M, Staudt H, Johansson P, Wachtveitl J, Oesterhelt D. 2009. Dodecin is the key player in flavin homeostasis of archaea. *J Biol Chem* **284:** 13068–13076.

Grininger M, Zeth K, Oesterhelt D. 2006. Dodecins: a family of lumichrome binding proteins. *J Mol Biol* **357:** 842–857.

Grkovic S, Brown MH, Roberts NJ, Paulsen IT, Skurray RA. 1998. QacR is a repressor protein that regulates expression of the *Staphylococcus aureus* multidrug efflux pump QacA. *J Biol Chem* **273:** 18665–18673.

Gusarov I, Nudler E. 1999. The mechanism of intrinsic transcription termination. *Mol Cell* **3:** 495-504.

Gutiérrez-Preciado A, Torres AG, Merino E, Bonomi HR, Goldbaum FA, García-angulo VA. 2015. Extensive identification of bacterial riboflavin transporters and their distribution across bacterial species. *PLoS ONE* **10:** e0126124.

Henkin TM. 2008. Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev* **22:** 3383-3390.

Howe JA, Wang H, Fischmann TO, Balibar CJ, Xiao L, Galgoci AM, Malinverni JC, Todd Mayhood, Villafania A, Nahvi A, et al. 2015. Selective small-molecule inhibition of an RNA structural element. *Nature* **526:** 672-677.

Kellenberger CA, Wilson SC, Hickey SF, Gonzalex TL, Su Y, Halberg ZF, Brewer TF, Lavarnoe AT, Carolson HK, Hsieh YF, Hammond MC. 2015. GEMM-I riboswitches from Geobacter sense the bacterial second messenger cyclic AMP-GMP. *Proc Natl Acad Sci USA* **112:** 5383–5388.

Kermani AA, Macdonald CB, Gundepudi R, Stockbridge RB. 2018. Guanidinium export is the primal function of SMR family transporters. *Proc Natl Acad Sci USA* **115:**3060-3065.

Kim JN, Roth A, Breaker RR. 2007. Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2'-deoxyguanosine. *Proc Natl Acad Sci USA* **104:** 16092–16097.

Kim JN, Blount KF, Puskarz I, Lim J, Link KH, Breaker RR. 2009. Design and antimicrobial action of purine analogues that bind Guanine riboswitches. *ACS Chem Biol* **4:** 915-927.

Landick R, Wang D, Chan CL. 1996. Quantitative analysis of transcriptional pausing by *Escherichia coli* RNA polymerase: his leader pause site as paradigm. *Methods Enzymol* **274:** 334–353.

Lewis K. 2000. Translocases: a bacterial tunnel for drugs and proteins. *Curr Biol* **10:** R678–R681.

Linster CL, Van schaftingen E, Hanson AD. 2013. Metabolite damage and its repair or pre-emption. *Nat Chem Biol* **9:** 72–80.

Lomovskaya O, Lewis K. 1992. Emr, an *Escherichia coli* locus for multidrug resistance. *Proc Natl Acad Sci USA* **89:** 8938–8942.

McCown PJ, Corbino KA, Stav S, Sherlock ME, Breaker RR. 2017. Riboswitch diversity and distribution. *RNA* **23:** 995–1011.

Mirihana Arachchilage G, Sherlock ME, Weinberg Z, Breaker RR. 2018. SAM-VI RNAs selectively bind *S*-adenosylmethionine and exhibit similarities to SAM-III riboswitches. *RNA Biol* **15:** 371-378.

Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E. 2002. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111:** 747–756.

Morita Y, Kodama K, Shiota S, Mine T, Kataoka A, Mizushima T, Tsuchiya T. 1998. NorM, a putative multidrug efflux protein, of *Vibrio parahaemolyticus* and its homolog in *Escherichia coli*. *Antimicrob Agents Chemother* **42:** 1778–1782.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29:** 2933-2935.

Nawrocki EP, Burge SW, Bateman A, Daub J, Eberhardt RY, Eddy SR, Floden EW, Gardner PP, Jones TA, Tate J, et al. 2015. Rfam 12.0: updates to the RNA families database. *Nucleic Acids Res* **43:** D130–137.

Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB, Breaker RR. 2017. Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Mol Cell* **65:** 220–230.

Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX, Breaker RR. 2013. Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat Chem Biol* **9:** 834–839.

Nelson JW, Sudarsan N, Phillips GE, Stav S, Lünse CE, McCown PJ, Breaker RR. 2015a. Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proc Natl Acad Sci USA* **112:** 5389–5394.

Nelson, J.W., Plummer, M.S., Blount, K.F., Ames, T.D., and Breaker, R.R. 2015b. Small molecule fluoride toxicity agonists. *Chem Biol* **22:** 527–534.

Ren A, Rajashankar KR, Patel DJ. 2012. Fluoride ion encapsulation by Mg^{2+} ions and phosphates in a fluoride riboswitch. *Nature* **486:** 85–89.

Robinson AE, Thomas NE, Morrison EA, Balthazor BM, Henzler-wildman KA. 2017. New free-exchange model of EmrE transport. *Proc Natl Acad Sci USA* **114:** E10083–E10091.

Serganov A, Huang L, Patel DJ. 2009. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* **458:** 233–237.

Serganov A, Nudler E. 2013. A decade of riboswitches. *Cell* **152:** 17–24.

Sherlock ME, Malkowski SN, Breaker RR. 2017a. Biochemical validation of a second guanidine riboswitch class in bacteria. *Biochemistry* **56:** 352–358.

Sherlock ME, Breaker RR. 2017b. Biochemical validation of a third guanidine riboswitch class in bacteria. *Biochemistry* **56:** 359–363.

Sherlock ME, Sudarsan N, Breaker RR. 2018a. Riboswitches for the alarmone ppGpp expand the collection of RNA-based signaling systems. *Proc Natl Acad Sci USA* **115:**6052-6057.

Sherlock ME, Sudarsan N, Stav S, Breaker RR. 2018b. Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. *eLife* **7:**e33908.

Soukup GA, Breaker RR. 1999. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5:** 1308–1325.

Stahmann KP, Arst HN, Althöfer H, Revuelta JL, Monschau N, Schlüpen C, Gätgens C, Wiesenburg A, Schlösser T. 2001. Riboflavin, overproduced during sporulation of Ashbya gossypii, protects its hyaline spores against ultraviolet light. *Environ Microbiol* **3:** 545-550.

Stockbridge RB, Robertson JL, Kolmakova-Partensky L, Miller C. 2013. A family of fluoride-specific ion channels with dual-topology architecture. *eLife* **2:** e01084.

Vidal-Aroca F, Giannattasio M, Brunelli E, Vezzoli A, Plevani P, Muzi-Falconi M, Bertoni G. 2006. One-step high-throughput assay for quantitative detection of β-galactosidase activity in intact Gram-negative bacteria, yeast, and mammalian cells. *Biotechniques* **40:** 433–440.

Vogl C, Grill S, Schilling O, Stülke J, Mack M, Stolz J. 2007. Characterization of riboflavin (vitamin B2) transport proteins from *Bacillus subtilis* and *Corynebacterium glutamicum*. *J Bacteriol* **189:** 7367–7375.

Wang JX, Lee ER, Morales DR, Lim J, Breaker RR. 2008. Riboswitches that sense *S*-adenosylhomocysteine and activate genes involved in coenzyme recycling. *Mol Cell* **29:** 691–702.

Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, 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.

Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR. 2010. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol* **11:** R31.

Weinberg Z, Breaker RR. 2011. R2R–software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* **12:** 3.

Weinberg Z, Lünse CE, Corbino KA, Ames, TD, Nelson JW, Roth A, Perkins KR, Sherlock ME, Breaker RR. 2017a. Detection of 224 candidate structured RNAs by comparative analysis of specific subsets of intergenic regions. *Nucleic Acids Res* **18:** 10811–10823.

Weinberg Z, Nelson JW, Lünse CE, Sherlock ME, Breaker RR. 2017b. Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. *Proc Natl Acad Sci USA* **114:** E2077–E2085.

West DW, Owen EC. 1973. Degradation of riboflavin by alimentary bacteria of the ruminant and man: production of 7,8-dimethyl-10-carboxymethylisoalloxazine. *Br J Nutr* **29:** 33–41.

Winkler WC, Cohen-Chalamish S, Breaker RR. 2002. An mRNA structure that controls gene expression by binding FMN. *Proc Natl Acad Sci USA* **99:** 15908–15913.

Yarnell WS, Roberts JW. 1999. Mechanism of intrinsic transcription termination and antitermination. *Science* **284:** 611-615.

Figures and Tables

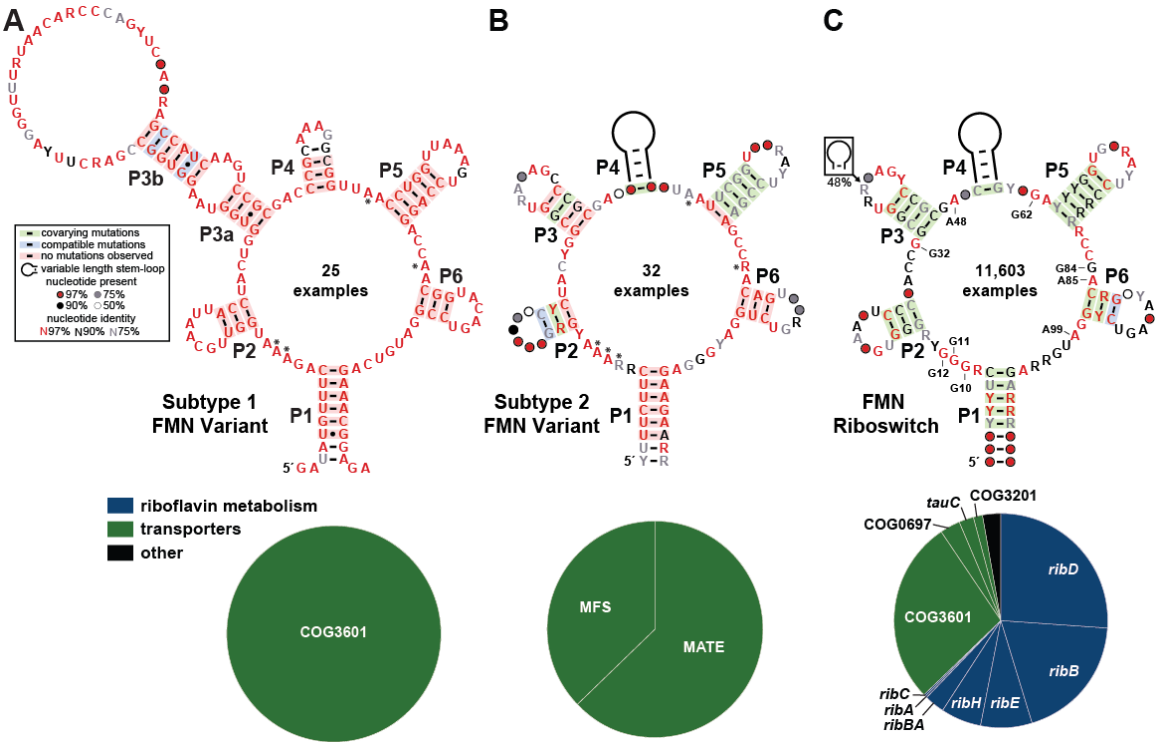

Figure 2-1. Variant FMN motif RNAs differ from the FMN riboswitch consensus and its typical genetic associations. (A) Consensus sequence and secondary structure model of 25 unique examples of subtype 1 FMN variant RNAs. Asterisks identify key nucleotides that differ from FMN riboswitch consensus. The key (box) describes the annotations in the consensus models. Bottom: Pie chart of the genes located immediately downstream of subtype 1 RNAs. (B) Consensus model and gene associations for 32 unique examples of subtype 2 RNAs. Additional annotations are as described for A. (C) Consensus model and gene associations for 11603 unique examples of FMN riboswitch aptamers. Certain nucleotides that are most directly involved in forming the ligand binding site for FMN are numbered according to the crystal structure model (PDB ID code 3F2Q) published previously (Serganov et al. 2009).

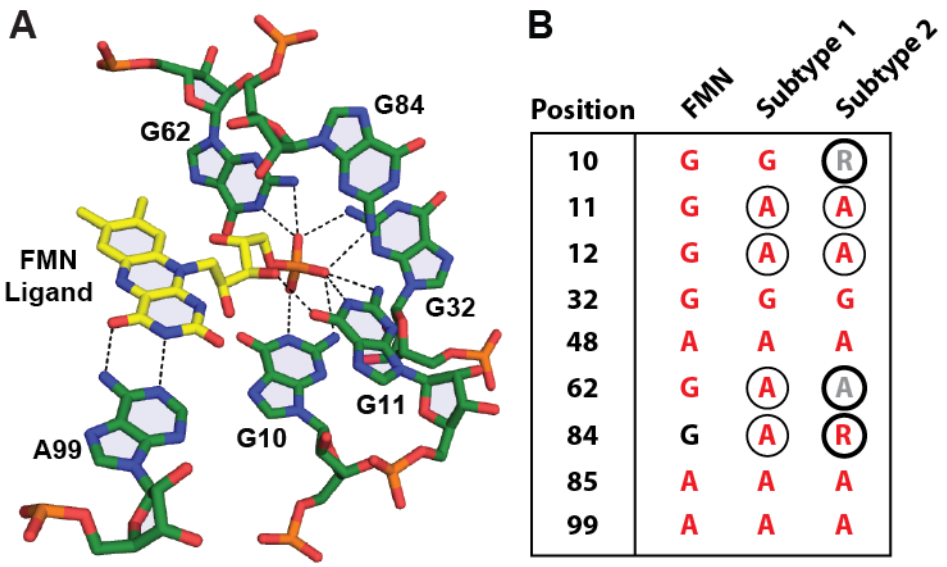

Figure 2-2. Atomic-resolution structural model for the binding site of an FMN riboswitch from *F. nucleatum*. (A) Atomic-resolution model of the ligand-binding site of an FMN riboswitch aptamer bound to FMN (Serganov et al. 2009) (PDB ID code 3F2Q). Dashed lines represent hydrogen-bonding contacts between aptamer nucleotides (numbered as depicted in **Figure 2-1C**) and the ligand, FMN. (*B*) Comparison of key equivalent nucleotide positions among FMN aptamer, subtype 1, and subtype 2 RNAs. Nucleotides are depicted according to their level of sequence conservation as described in the key to **Figure 2-1A**. Circled nucleotides are distinct from the consensus model for FMN riboswitches. Thick-lined circles identify nucleotides in the consensus for subtype 2 RNAs that differ from the consensus for subtype 1 RNAs. Nucleotide positions are depicted in A, except for positions 12, 48 and 85, which were omitted for image clarity.

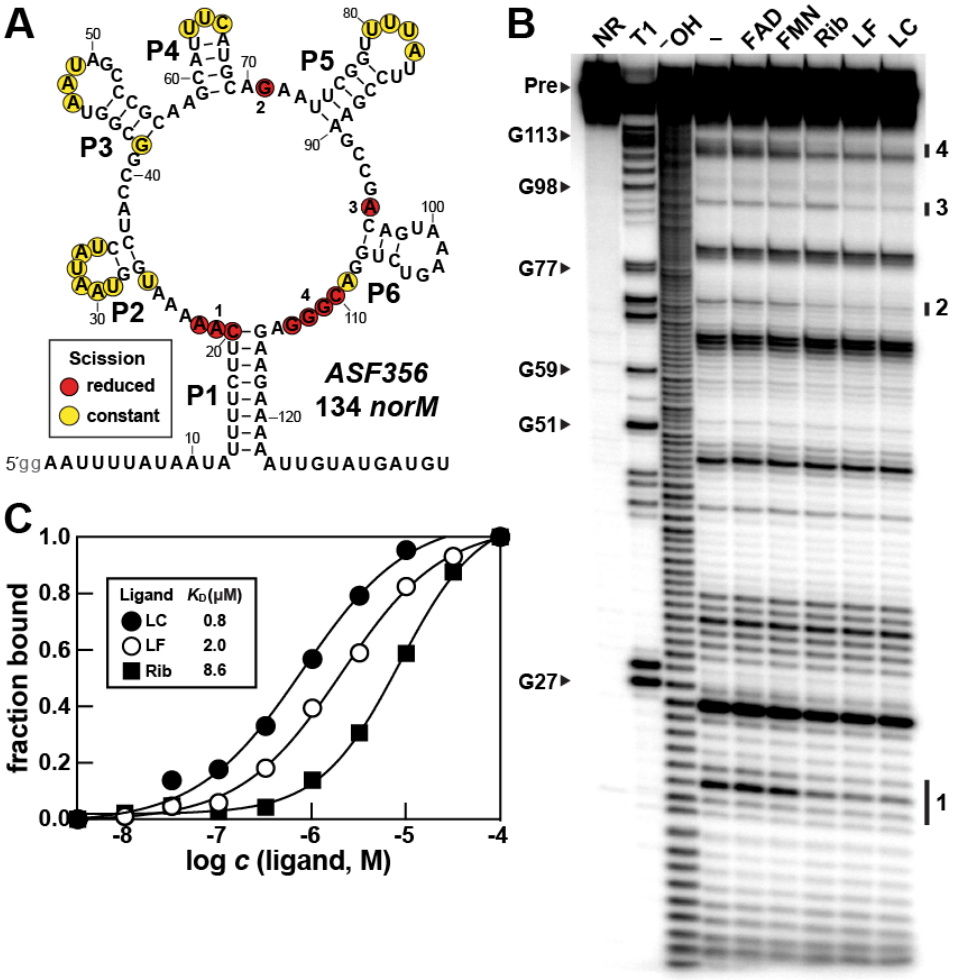

Figure 2-3. A representative of the subtype 2 variant FMN motif RNAs recognize FMN degradation products. (A) Sequence and secondary structure model of the 134 *norM* RNA construct. Lowercase letters on the 5′ terminus represent guanosine nucleotides added to the DNA template to promote transcription by T7 RNA polymerase. Regions of constant and reduced spontaneous RNA cleavage upon addition of ligand are indicated by yellow and red circles, respectively. This data is derived from the in-line probing analysis depicted in B. (B) PAGE analysis of in-line probing assays with 5′ ^{32}P-radiolabeled 134 *norM* RNA in the absence (–) or presence of FMN, riboflavin (Rib), lumiflavin (LF), or lumichrome (LC) at 100 μM. NR, T1, and ⁻OH designate RNAs that have undergone no reaction, that have been partially digested with RNase T1, or that

have been partially digested with alkali, respectively. Bands corresponding to certain RNase T1 digestion products (scission after G residues) are labeled according to the numbering system in A. Groups of bands denoted 1 through 4 identify regions of the RNA that undergo substantial change in response to ligand addition. (C) Plot of the fraction of RNA bound to the ligand versus the logarithm of the molar concentration (*c*) of the ligand. Fraction bound values were estimated based on the extent of band intensity changes at region 1 from data presented in **Figure 2-7**.

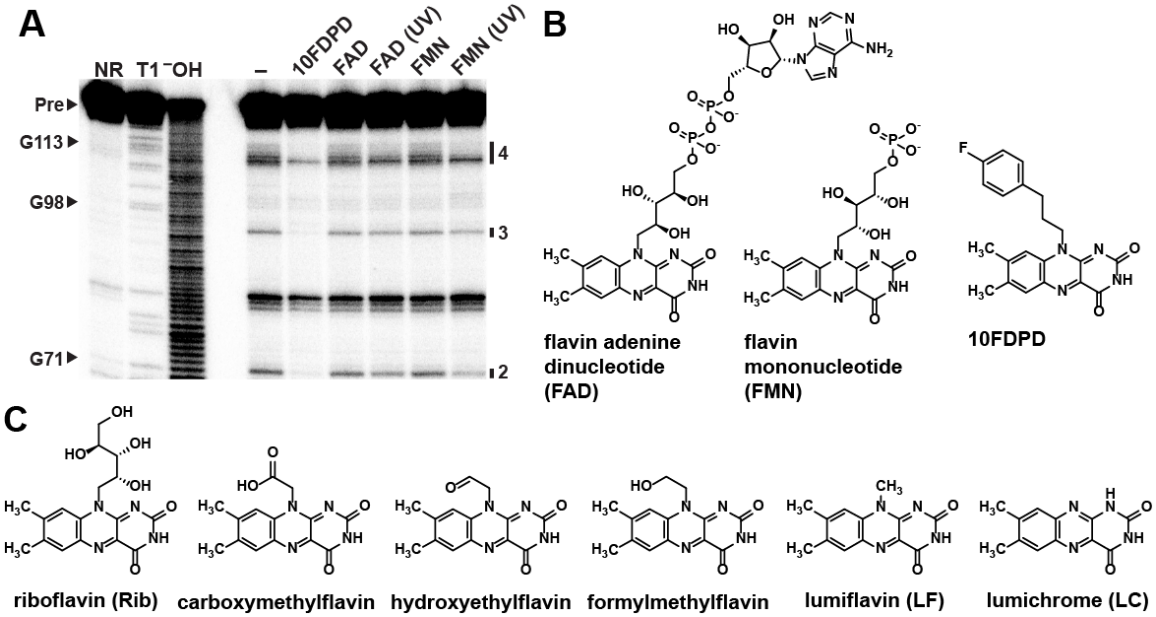

Figure 2-4. UV-treated FMN and FAD samples modulate the structure of a subtype 2 variant FMN riboswitch aptamer. (A) PAGE analysis of in-line probing assays of 5′ ^{32}P-radiolabeled 134 *norM* RNA from *C. sp. ASF356* (**Figure 2-3A**) in the absence (–) or presence of candidate ligands. Additional annotations are as described in the legend for **Figure 2-3B**. (B) Chemical structures of FAD, FMN and 10FDPD evaluated for binding in A. (C) Chemical structures of predicted FMN degradation products.

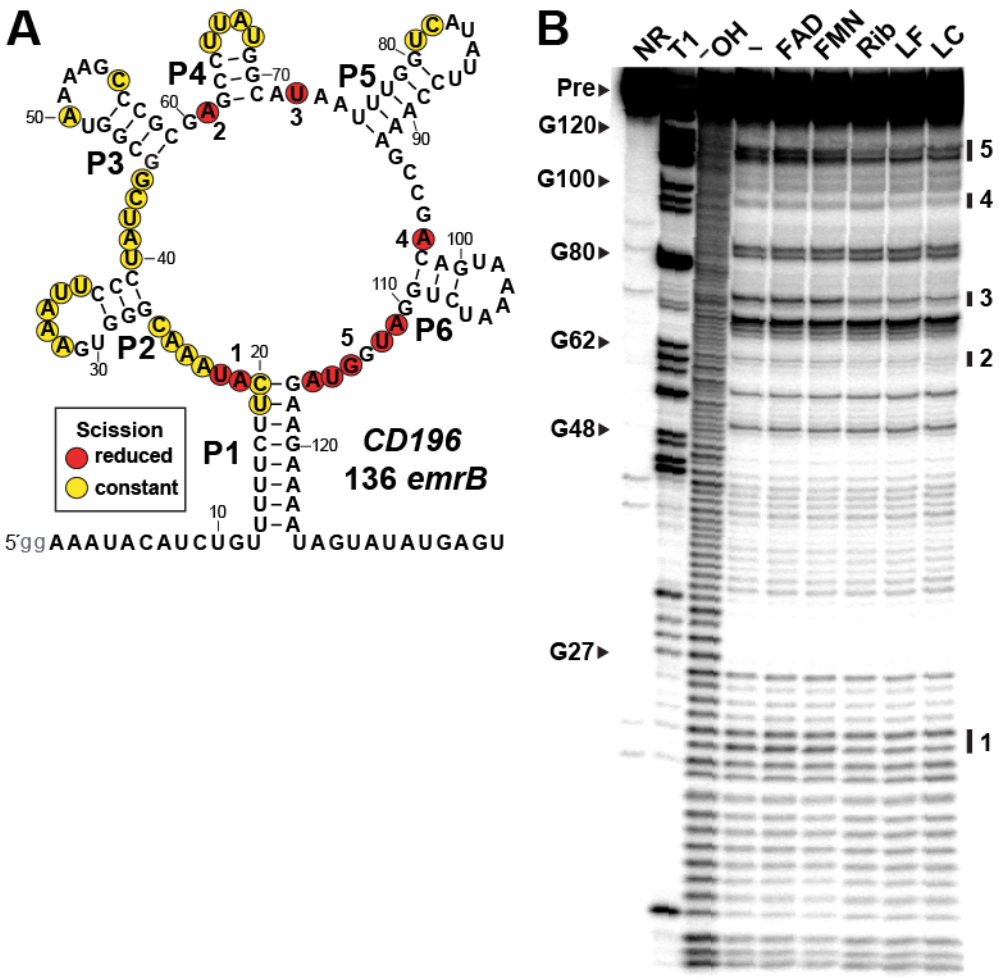

Figure 2-5. An additional representative of a subtype 2 variant FMN motif RNAs recognizes FMN derivatives. (A) Sequence and secondary structure of the 136 RNA derived from the *emrB* gene from *C. difficile* CD196. (B) PAGE analysis of in-line probing assays of 5′ ^{32}P-radiolabeled CD196 136 *emrB* RNA in the absence (–) or presence of various flavin compounds at 100 μM. Additional annotations are as described in the legend for **Figure 2-3**.

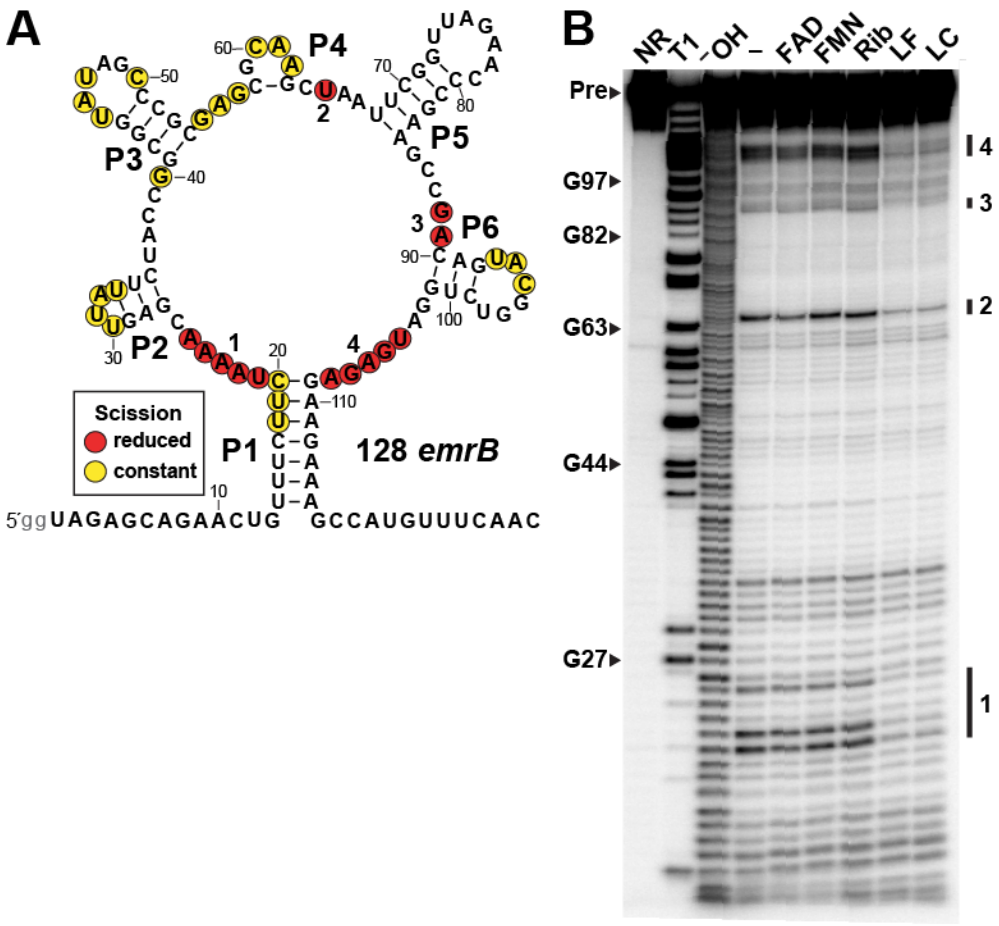

Figure 2-6. An additional representative of a subtype 2 variant FMN motif RNAs recognizes FMN derivatives. (A) Sequence and secondary structure model of the 128 RNA derived from the *emrB* gene from a metagenomic dataset sequence. (B) PAGE analysis of in-line probing assays of 5′ ^{32}P-radiolabeled 128 *emrB* RNA in the absence (–) or presence of various FMN analogs at 100 μM. Additional annotations are as described in the legend for **Figure 2-3**.

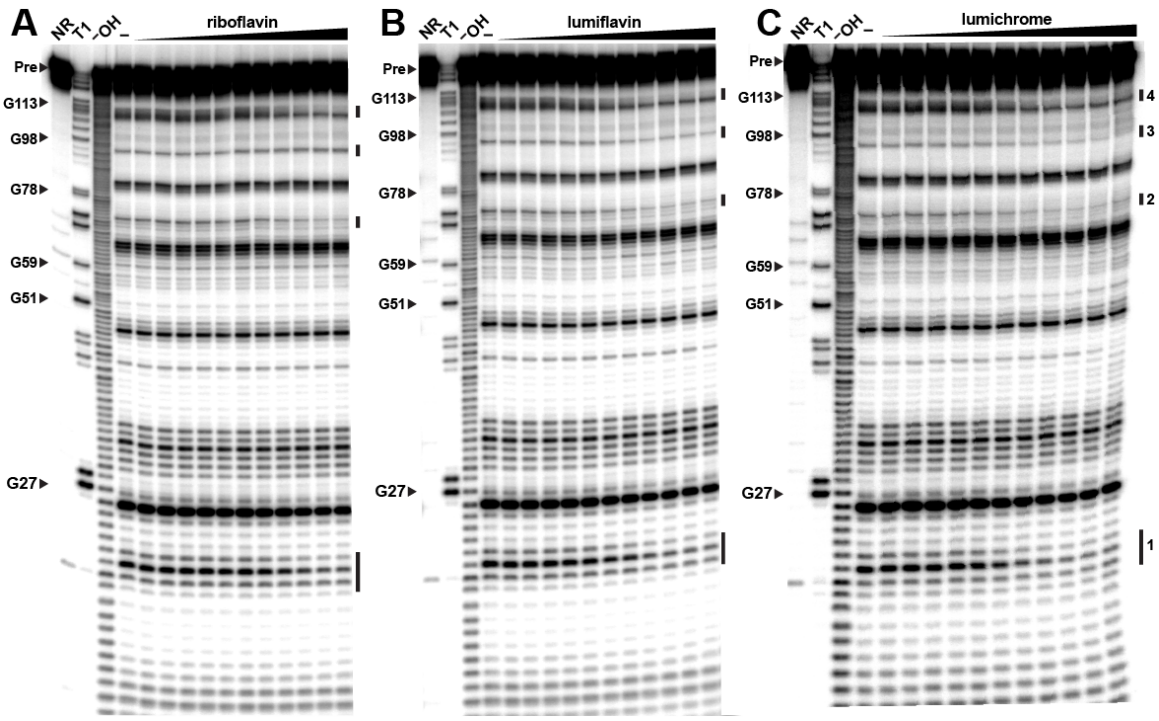

Figure 2-7. In-line probing analysis with 134 *norM* RNA. (A) PAGE analysis of in-line probing reactions in the absence (–) or presence of riboflavin ranging in concentration from 3.2 nM to 320 μM. (B) PAGE analysis of in-line probing reactions in the absence (–) or presence of lumiflavin ranging in concentration from 3.2 nM to 320 μM. (C) PAGE analysis of in-line probing reactions in the absence (–) or presence of lumichrome ranging in concentration from 3.2 nM to 320 μM. Additional annotation are as described in the legend for **Figure 2-3B**.

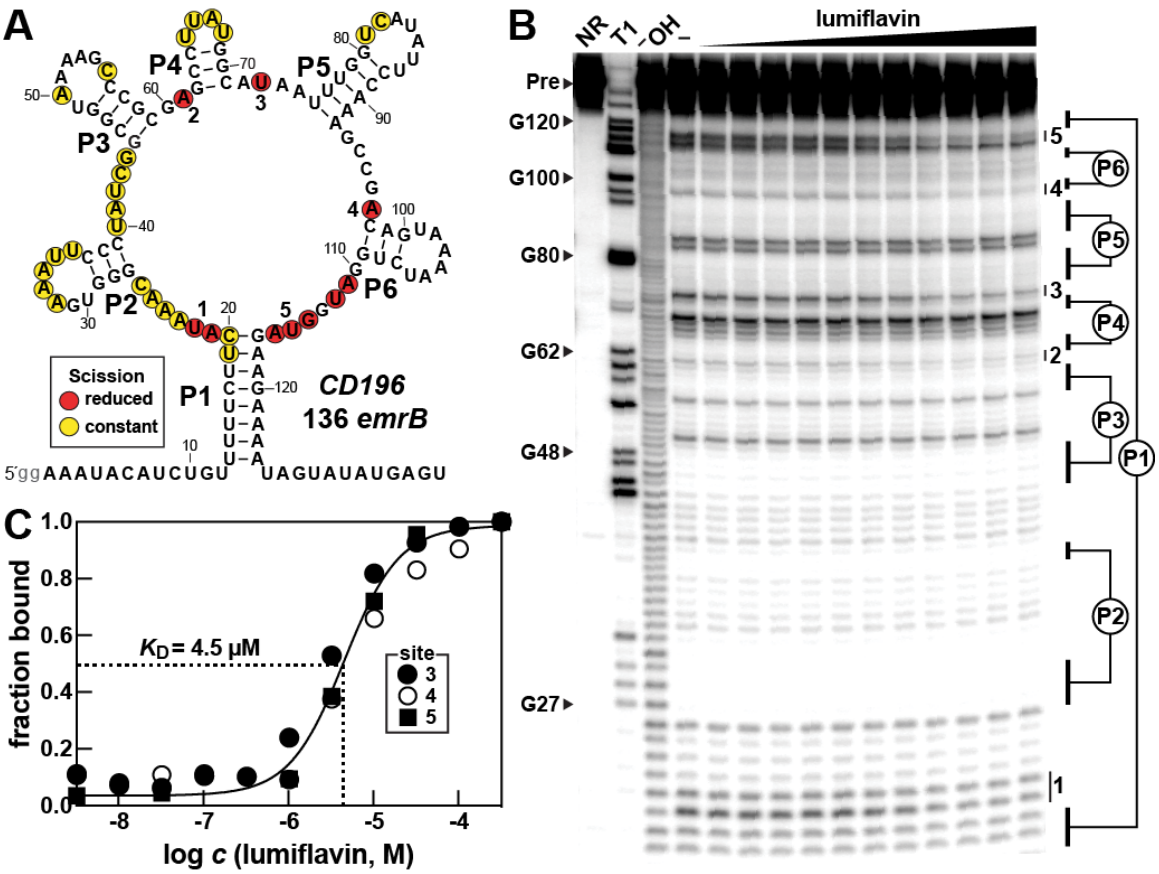

Figure 2-8. Another representative subtype 2 FMN variant RNA directly binds lumiflavin. (A) Sequence and secondary structure model of the 136 RNA derived from the *emrB* gene from *C. difficile CD196* depicting the sites of structural modulation by lumiflavin. (B) PAGE analysis of in-line probing reactions conducted in the absence (–) or presence of lumiflavin ranging in concentration from 3.2 nM to 320 µM. (C) The plot of the fraction of RNA bound to lumiflavin versus the logarithm of the molar concentration (*c*) of lumiflavin. Fraction bound values were determined based on the extent of band intensity changes at region 3, 4, and 5. Additional annotations are as described for **Figure 2-3**.

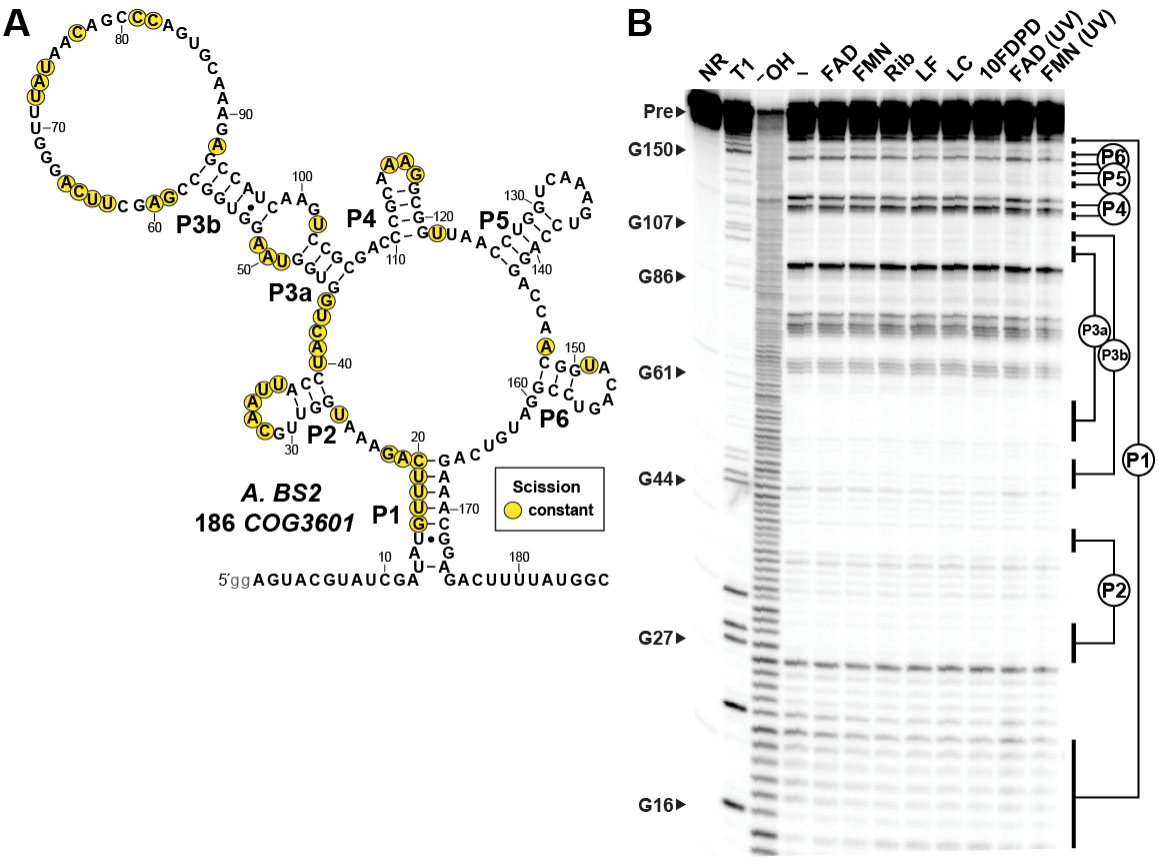

Figure 2-9. A subtype 1 variant FMN RNA rejects FMN and various other flavin derivatives. (A) Sequence and secondary structure of the 186 RNA derived from the gene annotated as *COG3601* from *Atopobium sp. BS2* (B) PAGE analysis of in-line probing assays of 5′ ^{32}P-radiolabeled 186 *COG3601* RNA in the absence (–) or presence of FMN or various derivative compounds at 100 μM. Additional annotation are as described in the legend for **Figure 2-3**.

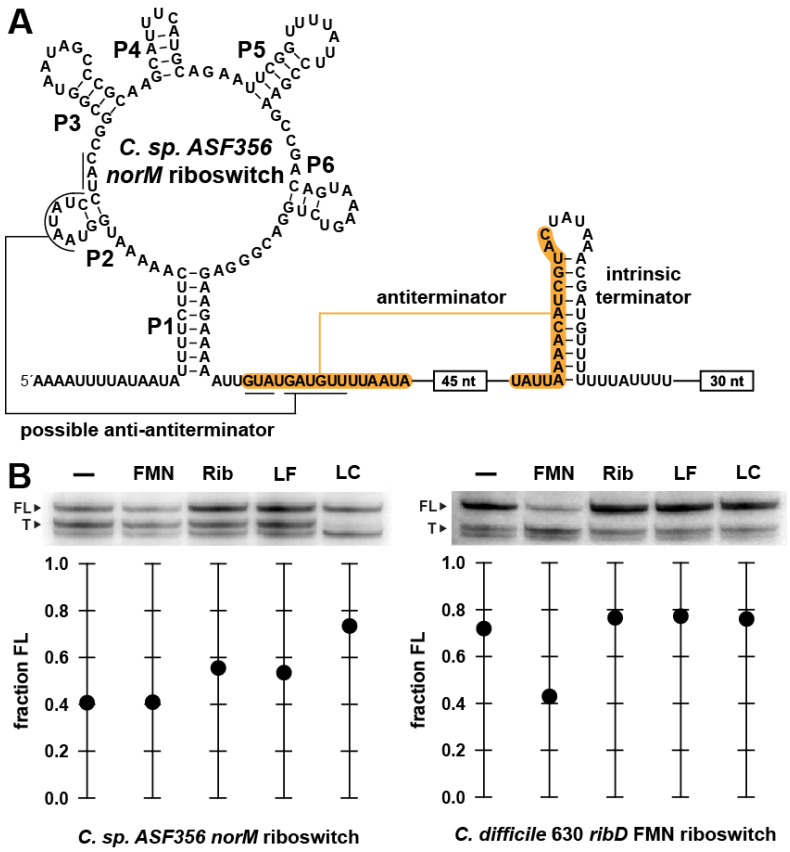

Figure 2-10. Subtype 2 and FMN riboswitch RNAs regulate transcription termination in response to different ligands. (A) Sequence and secondary structure model of the subtype 2 variant FMN riboswitch construct derived from the *norM* gene of *C. sp. ASF356* used for transcription termination assays. Terminated transcripts end within the U-rich tract immediately following the intrinsic terminator stem, whereas full-length transcripts from the DNA template used for this assay include the 30 nucleotides following the U-rich tract. (B) Top: PAGE analysis of single-round transcription termination assays conducted in the absence of ligand (–), or in the presence of 1 mM FMN, riboflavin (Rib), lumiflavin (LF), or lumichrome (LC). FL and T denote full-length and terminated RNA transcripts, respectively. Bottom: Values for the fraction of FL RNA generated are derived from the PAGE data presented.

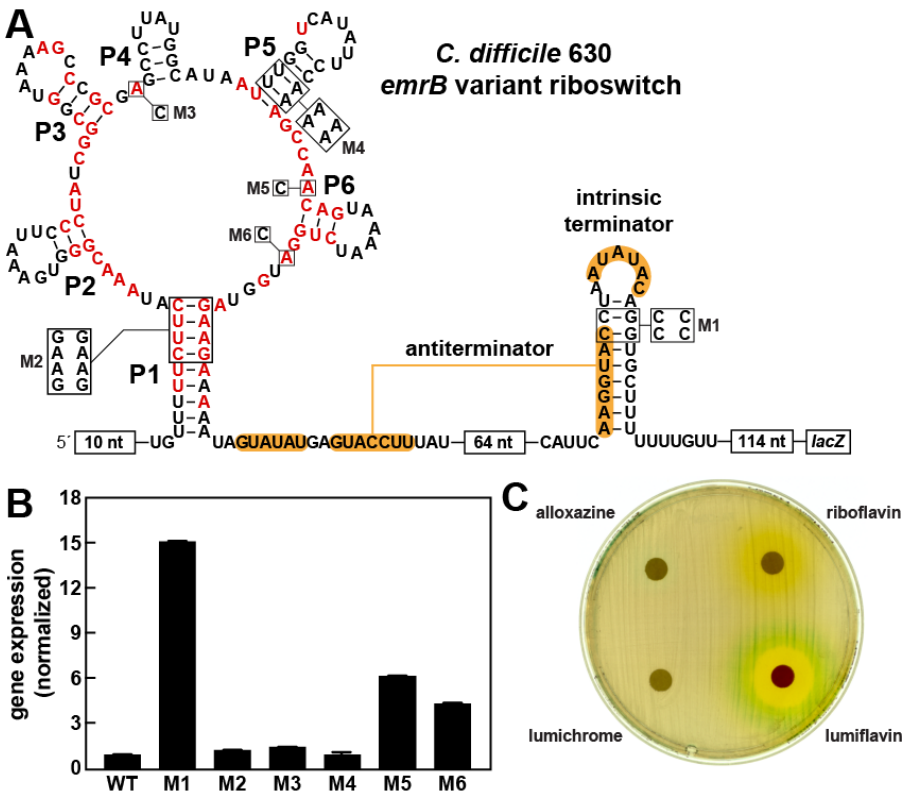

Figure 2-11. In vivo assessment of gene control by a subtype 2 variant RNA from *C. difficile*. (A) Sequence and secondary structure of the regulatory region of the riboswitch-reporter construct by joining the variant FMN riboswitch RNA from the *emrB* gene of *C. difficile* 630 to a *lacZ* gene. Red nucleotides are >97% conserved. Mutation of conserved nucleotides and secondary structures are designated M1 through M6. (B) Plot of the normalized β-galactosidase reporter gene expression levels of *B. subtilis* cells containing WT and various mutant reporter constructs in rich medium. Error bars indicate the standard deviation from three independent experiments. (C) Agar diffusion assays were conducted with LB agar plates supplemented with 100 μg mL^{-1} X-gal and inoculated with *B. subtilis* cells carrying the WT reporter construct as described in A. Filter disks were infused with 10 μL of alloxazine (50 mM), riboflavin (35 mM), lumichrome (10 mM), or lumiflavin (80 mM).

Table 2-1. Sequences of synthetic DNAs used in Chapter 2.

Name	Sequence	Purpose
FMN001	TAATACGACTCACTATAGGAATTTTA TAATATTTTCTTCAAAAATGGTAATA TCCTACCGGCGGTAATAGCCCGCAA GCATTTCATGC	Forward primer for the construction of a template for transcription of the *C. ASF356* 134 *norM* RNA
FMN002	ACATCATACAATTTTTCTTCTCCCG TCCAGACTTTACTGTCGGCTTCGG AATAAAACCGAATTCTGCATGAAA TGCTTGCGGGC	Reverse primer for the construction of a template for transcription of the *C. ASF356* 134 *norM* RNA
FMN003	TAATACGACTCACTATAGGAAATAC ATCTGTTTTTCTTCATAAACGGGTGA AATTCCCTATCGGCGGTAAAAGCCC GCGAGCCTTATGGC	Forward primer for the construction of a template for transcription of the *C. difficile CD196* 136 *emrB* RNA
FMN004	ACTCATATACTATTTTCTTCTACCATC CAGATTTTACTGTCGGCTTTGGAATA TGACCAAATTATGCCATAAGGCTCGC GGGCTTTTACC	Reverse primer for the construction of a template for transcription of the *C. difficile CD196* 136 *emrB* RNA
FMN005	TAATACGACTCACTATAGGTAGAGCA GAACTGTTTCTTCTAAAACGAGTTATT CTACCGGCGGTATAGCCCGCGAGCGC AAGC	Forward primer for the construction of a template for transcription of the 128 *emrB* RNA
FMN006	GTTGAAACATGGCTTTCTTCTCTCATC CAGACCGTACTGTCGGCTTCGGGTTCT AACCGAATTAGCTTGCGCTCGCGGGCT ATACC	Reverse primer for the construction of a template for transcription of the 128 *emrB* RNA
FMN007	TAATACGACTCACTATAGGAGTACGTA TCGATATGTTTCAGAAATGGTTGCAAT TACCTACTGGTGGTAAGGTGGCCGAGC TTCAGGGTTTATAACAGCCCAGTGCAA AGAGC	Forward primer for the construction of a template for transcription of the *C. BS2* 186 *COG3601* RNA
FMN008	GCCATAAAAGTCTCCGTTTCTGACATCC GGACTGTACCGTTGGTCCTGGACTTTAA CCAGGTTAACCGCCTTTGCGGGTCGCG GACTTGATGGCTCTTTGCACTGGGCTG	Reverse primer for the construction of a template for transcription of the *C. BS2* 186 *COG3601* RNA

FMN009	CAAGCGCAGCATAATGTTGATCCTTT TAAATAAGTCTGATAAAATGTGAACT AAAATTTTATAATATTTTCTTCAAAAA TGGTAATATCCTACCGGCGGTAATAG CCCGCAAGCATTTCATGCAGAATTCGG TTTTATTCCGAAGCCGACAGTAAAGTC TGGACGGGAGAAGAAAAATTGTATGA TGTTTTAATATCAATTATTTTTCTTCTC TTTCATATAGGTTATAGAAGAAAATTCT ATTAAAACATCGTACTATAAACGATGTT TTTTTATTTTGTAAAAAAGAAAAGAAAG GAAGTGGGATCC	IDT G-block containing the *C. ASF356* variant FMN riboswitch controlled by the *B. subtilis lysC* promoter for single-round transcription termination
FMN010	CAAGCGCAGCATAATGTTGATCC	Forward primer for the amplification of *C. ASF356* variant FMN riboswitch termination construct
FMN011	GGATCCCACTTCCTTTCTTTTC	Reverse primer for the amplification of *C. ASF356* variant FMN riboswitch termination construct
FMN012	TACGACGAATTCCAAAAATAATGTTG ATCCTTTTAAATAAGTCTGATAAAATG TGAACTAATTTAATACTTAATCTTCGG GGTAGGG	Forward primer for the amplification of FMN *ribD* riboswitch single-round transcription termination template from *C. difficile* 630 genomic DNA genomic DNA
FMN013	TCCTTTGGATCCAATATGTACACCTCC TGTTTAATTGG	Reverse primer for the amplification of FMN ribD riboswitch single-round transcription termination template from *C. difficile* 630 genomic DNA genomic DNA
FMN014	TACGACGAATTCCAAAAATAATGTTGA TCCTTTTAAATAAGTCTGATAAAATGTG AACTAAAATACATCTGTTTTTCTTCATA AACGGGTG	Forward primer for the construction of *C. difficile* 630 variant FMN reporter construct containing the *B. subtilis lysC*

		promoter containing sites for cloning into pDG1661
FMN015	TCCTTTGGATCCGATAACACTCCAATCT TAATAGATTTTATA	Reverse primer for the construction of *C. difficile* 630 variant FMN reporter construct construct containing the *B. subtilis lysC* promoter containing sites for cloning into pDG1661
FMN016	GTGAACTAAAATACATCTGTTTTTGAAG ATAAACGGGTGAAATTCCCTATCG	Forward primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC* promoter containing M1 via QuikChange mutagenesis
FMN017	CGATAGGGAATTTCACCCGTTTATCTTCAA AAACAGATGTATTTTAGTTCAC	Reverse primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC* promoter containing M1 via QuikChange mutagenesis
FMN018	CGGCGGTAAAAGCCCGCGCGCCTTATG GCATAATTTG	Forward primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC* promoter containing M2 via QuikChange mutagenesis
FMN019	CAAATTATGCCATAAGGCGCGCGGGCT TTTACCGCCG	Reverse primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC* promoter containing M2 via QuikChange mutagenesis
FMN020	CGCGAGCCTTATGGCATAATAAGGTCA TATTCCAAAGCCAAC	Forward primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC* promoter containing M3 via QuikChange mutagenesis
FMN021	GTTGGCTTTGGAATATGACCTTATTATG CCATAAGGCTCGCG	Reverse primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC*

		promoter containing M3 via QuikChange mutagenesis
FMN022	GTCATATTCCAAAGCCACCAGTAAAAT CTGGATGG	Forward primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC* promoter containing M4 via QuikChange mutagenesis
FMN023	CCATCCAGATTTTACTGGTGGCTTTGGA ATATGAC	Reverse primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC* promoter containing M4 via QuikChange mutagenesis
FMN024	CAACAGTAAAATCTGGCTGGTAGAAGA AAATAG	Forward primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC* promoter containing M5 via QuikChange mutagenesis
FMN025	CTATTTTCTTCTACCAGCCAGATTTTA CTGTTG	Reverse primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC* promoter containing M5 via QuikChange mutagenesis
FMN026	GTACCTAATATACACCTGCTTTTTTTGTTG	Forward primer for construction of the *C. difficile* 630 variant FMN reporter construct with the *lysC* promoter containing M6 via QuikChange mutagenesis

Chapter Three

Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class

Largely adapted from the following publication:

Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB, Breaker RR. 2017. Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Mol Cell* **65:** 220-230.

Summary

The guanidyl moiety is a component of fundamental metabolites including the amino acid arginine, the energy carrier creatine, and the nucleobase guanine. Curiously, reports regarding the importance of free guanidine in biology are sparse and no biological receptors that specifically recognize this compound have been previously identified. In this chapter, my colleagues and I demonstrate that many members of the *ykkC* motif RNA, the longest-unresolved riboswitch candidate, naturally sense and respond to guanidine. This RNA is found throughout much of the bacterial domain of life, where it commonly controls the expression of proteins annotated as urea carboxylases and multidrug efflux pumps. Analyses performed by my colleagues and I reveal that these proteins likely function as guanidine carboxylases and guanidine transporters, respectively. Furthermore, we demonstrate that bacteria are capable of endogenously producing guanidine. These and related findings demonstrate that free guanidine is a biologically relevant compound and several gene families exists that can alleviate guanidine toxicity.

Introduction

Guanidine was first discovered as a thermal decomposition product of the nucleobase guanine (Strecker, 1861). Industrial-scale production of guanidine has been exploited for the manufacture of explosives and other chemicals, but this compound is perhaps best known to the scientific community as a commonly-used reagent for protein denaturation (Greene and Pace, 1974). Microbes capable of metabolizing guanidine have been previously studied, in part due to the interest in guanidine bioremediation (Bierema, 1909; Ebisuno and Takimoto, 1981; Iwanoff and Awetissowa, 1931; Kihara et al., 1955; Mitchell, 1985, 1987). Guanidine also has been proposed to be generated by human cells as part of the guanidine cycle via the cleavage of canavanine (Kihara et al., 1955; Natelson and Sherwin, 1979). Despite these scattered findings, guanidine has never been considered to be a major cellular metabolite. Indeed, no biological receptor has previously been identified that specifically recognizes guanidine.

In this chapter, my colleagues and I determined that many bacteria employ riboswitches (Breaker, 2011; Serganov and Nudler, 2013) to selectively bind guanidine (most likely as the guanidinium cation) and regulate genes encoding proteins that overcome its toxicity. Over the past fifteen years, more than 30 different riboswitch classes have been demonstrated to control a variety of metabolic or physiological processes, including previously underappreciated aspects of biology. Typically, widespread bacterial riboswitches recognize metabolites or signaling molecules that are themselves widespread and broadly important to life.

One of the most common uncharacterized RNA motifs identified in bacteria over the past two decades is the *ykkC* riboswitch candidate. This RNA motif was one of eight

classes discovered in an earlier study using comparative sequence analysis to identify riboswitches (Barrick et al., 2004). Seven of the motifs identified in this previous study have since been characterized, and six of these were found to control important aspects of biology related to cell wall biosynthesis (Winkler et al., 2004; Nelson et al., 2013) and sporulation (Marchais et al., 2011; Nelson et al., 2013), Mg^{2+} homeostasis (Dann et al., 2007), Mn^{2+} homeostasis (Dambach et al., 2015; Price et al., 2015), glycine metabolism (Mandal et al., 2004), and biosynthesis of modified nucleotides (Roth et al., 2007). Some of these RNA candidates were considered "orphan" riboswitches because their ligands had not immediately been validated experimentally (Block et al., 2010; Meyer et al., 2011). The *ykkC* motif RNA is the last orphan remaining from this initial collection of RNA classes (Barrick et al., 2004) to resist all previous efforts at experimental validation (Meyer et al., 2011).

Members of the *ykkC* class control an eclectic distribution of genes including those encoding predicted urea carboxylases and their associated proteins, as well as predicted allophanate hydrolases, arginases, creatininases, and nitrate/sulfate/bicarbonate transporters (Barrick et al., 2004; Meyer et al., 2011). The RNA is also commonly found upstream of genes annotated as *emrE* or *sugE*, which are predicted to encode small multidrug resistance (SMR) efflux pumps (Jack et al., 2000; Schuldiner, 2009). For example, the *Escherichia coli* EmrE protein has been demonstrated to be highly indiscriminate and is thought to have evolved to protect the cell against a wide variety of small molecule toxins (Schuldiner, 2012). In addition, some *ykkC* motif RNAs are found upstream of genes involved in branched-chain amino acid biosynthesis and purine metabolism (Meyer et al., 2011). Frequently, these RNAs are predicted to control the

entire *de novo* purine biosynthetic operon of their host organism. Due to the diversity of genes associated with this motif, no single compound emerged as a probable riboswitch ligand.

My colleagues hypothesized that *ykkC* RNAs might recognize some toxic ligand, as they are associated with disparate metabolic genes and genes that encode numerous types of transporter proteins. Accordingly, my colleague James Nelson screened a wide variety of growth conditions, including exposure to numerous toxic compounds, for those that trigger riboswitch activation in cells. Of the roughly 2,000 conditions tested, only the addition of guanidine hydrochloride led to riboswitch-mediated gene expression. My colleague Madeline Sherlock subsequently demonstrated that certain *ykkC* motif RNAs are capable of specifically recognizing guanidine *in vitro* at concentrations many orders of magnitude below those required for guanidine to non-specifically affect RNA structure. Furthermore, she performed biochemical analysis of a previously studied urea carboxylase (Kanamori et al., 2004) revealed that this protein prefers guanidine over urea as a substrate. Additionally, my colleague Randy Stockbridge examined one of the SugE transporters whose expression is controlled by the riboswitch and collected evidence that it selectively recognizes guanidine. Finally, I demonstrated that bacteria are capable of endogenously producing guanidine under laboratory conditions. These and other findings reveal that the majority of *ykkC* motif RNAs are members of a broadly distributed riboswitch class that specifically senses guanidine and controls the expression of a variety of genes that together constitute a previously unknown super-regulon for its metabolism and transport.

Results

The *ykkC* motif is a widespread orphan riboswitch candidate predicted to control the expression of a variety of transporter and metabolic genes

By performing a series of homology searches, James Nelson identified ~2000 distinct *ykkC* RNA sequences in Actinobacteria, Firmicutes, Proteobacteria and Cyanobacteria. Through additional experimental efforts as discussed below, Madeline Sherlock established that there are at least two distinct subtypes of *ykkC* motif RNAs. Most of these RNAs are of the predominant "subtype 1" (**Figure 3-1A**), whereas a rarer variant *ykkC* motif called "subtype 2" (**Figure 3-2**) carries prevalent alterations in three otherwise highly-conserved regions of the *ykkC* motif (**Figure 3-2A**). Analysis of the nucleotide sequences and structures located 3′ of this motif revealed the frequent presence of an adjoining expression platform. Therefore, we predicted that the vast majority of *ykkC* motif representatives of both subtypes function as riboswitches by activating expression of downstream genes upon recognition of the cognate ligand by, for instance, precluding formation of an intrinsic transcription terminator stem (Gusarov and Nudler 1999; Yarnell and Roberts, 1999).

The identities of the ligands recognized by newly discovered riboswitch classes frequently can be inferred based on the functions of proteins whose genes reside immediately downstream. Unfortunately, the total collection of genes associated with *ykkC* motif RNAs (**Figure 3-1B**, **Figure 3-2B**) did not suggest any likely ligand candidates in addition to those that already had been examined (Meyer et al., 2011). Intriguingly, two other bacterial RNA motifs, called mini-*ykkC* (Weinberg et al., 2007) and *ykkC*-III (Weinberg et al., 2010), are often found upstream of genes similar to those

associated with subtype 1 RNAs. This suggests that these three RNA motifs might all respond to the same ligand, either directly or by employing an intermediary protein factor. We therefore anticipated that any insight gained by the study of subtype 1 RNAs would also advance our understanding of the biological and biochemical functions of these other two RNA classes as further discussed below.

Guanidine triggers gene expression mediated by a subtype 1 *ykkC* motif RNA

James Nelson screened a large diversity of compounds and growth conditions for any factor that might result in activation of a riboswitch reporter construct. He used a reporter created previously (Meyer et al., 2011) by fusing the subtype 1 *ykkC* motif representative naturally associated with the *Bacillus subtilis ykkC* coding region to a β-galactosidase (*lacZ*) reporter gene (**Figure 3-3A**). Of the nearly 2,000 conditions examined, only the addition of guanidine triggered reporter gene expression mediated by the *ykkC* motif RNA (**Figure 3-4A**).

To experimentally confirm this hit, I grew the *B. subtilis* reporter strain with various concentrations of added guanidine (**Figure 3-4B**). Reporter gene expression progressively increases with increasing guanidine, beginning with concentrations as low as 100 μM. Since guanidine is an effective protein denaturant above 1 M (Greene and Pace, 1974), this suggests that guanidine's influence on reporter gene expression is unlikely to be due to nonspecific denaturation of a riboswitch or protein regulatory factor.

Guanidine-dependent expression also occurs if the native promoter for the *ykkC* gene (Meyer et al., 2011) is replaced with an unrelated promoter (**Figure 3-3B**), which indicates that the *ykkC* motif RNA is responsible for ligand-induced gene regulation.

Furthermore, disruption of conserved RNA sequence or structural features, such as in reporter constructs M1 and M2, results in a loss of expression (**Figure 3-3C**). In contrast, incorporation of mutations in construct M3 that retain a conserved RNA structure also permits guanidine-dependent regulation.

To further demonstrate that guanidine selectively triggers RNA-mediated gene expression, I created another reporter construct in *E. coli* cells based on a representative subtype 1 example associated with the *Klebsiella pneumoniae tauA* gene, which encodes a putative ATPase located near additional genes commonly associated with *ykkC* motif RNAs. Again, the addition of guanidine to the culture medium activates expression of the WT reporter construct, whereas mutation of a single conserved nucleotide in construct M4 disrupts regulation (**Figure 3-5A, B, C**). Furthermore, guanidine does not trigger expression of a member of an unrelated riboswitch class (**Figure 3-5D, E**). These results are consistent with the hypothesis that subtype 1 representatives of *ykkC* motif RNAs function as guanidine-dependent riboswitches, and hereafter these RNAs will be called guanidine class I or 'guanidine-I' riboswitches.

The guanidine-I riboswitch representative examined from *B. subtilis* controls the expression of an apparent two-gene operon comprised of the *ykkC* and *ykkD* coding regions. The proteins encoded by these genes are predicted to form a heterodimer and function as an SMR efflux pump (Jack et al., 2000). If guanidine is the biologically-relevant ligand for this riboswitch class, then this protein complex might function to alleviate guanidine toxicity by reducing its concentration in cells. Indeed, upon deleting the *ykkC* and *ykkD* genes, we observed decreases in both the concentrations of guanidine

required to inhibit bacterial growth (**Figure 3-3D**) and to activate riboswitch-mediated expression of *lacZ* (**Figures 3-3E, F**).

As noted earlier, guanidine is a well-known chaotropic agent, and therefore might adversely affect RNA structure in a manner similar to its protein denaturation mechanism. Alternatively, structural reorganization by ionic stabilization of the RNA could be mediated by guanidinium cations (**Figure 3-3B**), which represent the predominant form of guanidine at biologically-relevant pH conditions. Metal and nonmetal (e.g. amine and polyamine) cations are well known to interact non-specifically with RNA and induce structural changes (Hermann and Westhof, 1999; Pyle, 2002). Given these potential modes of non-specific interaction, we sought to provide several additional lines of evidence that guanidine is the biologically relevant ligand as described in the following sections.

Molecular recognition determinants for guanidine binding by a riboswitch

I examined the determinants of guanidine recognition by testing a number of chemically related compounds both for reporter gene regulation and for binding *in vitro*. Of the 16 guanidine analogs examined using the *B. subtilis* riboswitch reporter (**Figure 3-3A**), none resulted in expression equal to that induced by guanidine (**Figure 3-5F**). Only the closest chemical analogs of guanidine, specifically hydroxy-, amino-, and methylguanidine, yielded reporter gene expression above background. Of these, hydroxyguanidine is unstable, and yields guanidine upon degradation (Walker and Walker, 1959). Therefore, it is possible that the signal generated by this compound is not caused by the original analog. Strikingly, the addition of up to molar amounts of urea failed to yield any

measurable increase in reporter gene expression. These results suggest that the *ykkC* motif RNAs examined above have evolved to serve as highly selective sensors for free guanidine, and to exclude even closely-related analogs such as urea that might otherwise inappropriately trigger gene expression. It seems unlikely that the riboswitch might preferentially respond to another natural compound that incorporates a guanidyl moiety since any modification to a nitrogen atom causes a loss of affinity.

We considered the possibility that guanidine might induce cells to produce a structurally unrelated signaling compound, or perhaps a metabolic byproduct of guanidine degradation, that then directly triggers riboswitch function. To determine whether the riboswitch aptamer directly recognizes guanidine, Madeline Sherlock conducted in-line probing assays (Regulski and Breaker, 2008; Soukup and Breaker, 1999) to monitor RNA structural changes that are indicative of riboswitch modulation by ligand binding. Initially, she examined a 104 nucleotide aptamer called 104 *uca* (**Figure 3-6A**) that was derived from the guanidine-I riboswitch associated with the putative urea carboxylase gene from the bacterium *Nitrosomonas europaea*. Guanidine causes decreased levels of spontaneous RNA strand scission at two sites (**Figure 3-6B**), indicating that specific nucleotides within these conserved regions undergo structural reorganization. A dissociation constant (K_D) of ~60 μM for the guanidine-RNA complex is estimated from these data (**Figure 3-6C**). These results are typical of the structural changes that occur when representatives of other riboswitches classes are exposed to their cognate ligands. In contrast, 20 different compounds carrying a guanidyl moiety or a charged amine group were also examined by in-line probing (**Figure 3-7**), but only guanidine triggers changes in the shape of the RNA aptamer.

The structural changes observed with the 104 *uca* RNA are unlikely to be due to the chaotropic nature of guanidine. Spontaneous RNA strand scission is reduced by guanidine, which indicates that the RNA is becoming more rather than less structured. The latter effect would be expected if guanidine were functioning as a denaturant. Also, the K_D for RNA binding is more than five orders of magnitude lower than is necessary to cause protein denaturation (Greene and Pace, 1974). Moreover, ligand-mediated modulation occurs with a dose-response curve consistent with a 1-to-1 binding interaction (**Figure 3-6C**). This suggests that the mechanism of riboswitch function is due to the docking of a single ligand in a single saturable binding site, rather than by non-specific neutralization of negatively charged phosphate groups on the RNA backbone by guanidinium.

For other riboswitch classes, many highly-conserved nucleotides within the aptamer domain establish critical structural contacts to form a selective ligand binding pocket. Therefore, if guanidine is being selectively bound by the RNA, then the mutation of any single conserved nucleotide is likely to adversely affect ligand-induced riboswitch function. To test this hypothesis, Madeline Sherlock examined a larger collection of mutant 104 *uca* aptamers (**Figure 3-8**, constructs M8 through M16) that individually alter nine highly-conserved nucleotides. In each instance, the single mutation either reduces (M10 and M12) or eliminates (M8, M9, M11, M13-M16) guanidine binding. These results suggest that guanidine-I riboswitches exploit numerous conserved nucleotides to form a selective binding pocket for guanidine. Indeed, x-ray structural analysis of a guanidine-I riboswitch aptamer confirms the critical roles of all highly-conserved

nucleotides to form a 1-to-1 complex between RNA and guanidine (Reiss and Strobel, 2016).

Similar ligand-induced structural modulation of other guanidine-I riboswitch aptamers also is observed (**Figures 3-9, Figure 3-10**). In contrast, James Nelson and Madeline Sherlock showed that the addition of up to 1 mM guanidine to a subtype 2 *ykkC* motif RNA (data not shown) or 0.5 M guanidine to aptamers from four other riboswitch classes (azaaromatic, c-di-AMP, c-AMP-GMP, ZTP; data not shown) failed to induce structural changes. These results again indicate that guanidine, even at exceedingly high concentrations, does not cause RNA structural changes by non-specific interactions. Importantly, the addition of guanidine to representative mini-*ykkC* and *ykkC*-III motif RNAs also induces structure modulation (Sherlock and Breaker, 2017; Sherlock et al., 2017). These findings are consistent with the original hypothesis that mini-*ykkC* and *ykkC*-III motif RNAs might represent additional distinct riboswitch classes that respond to the same ligand as *ykkC* motif RNAs with similar gene associations (Weinberg et al., 2010).

Guanidine selectively induces riboswitch-mediated transcription

James Nelson assessed the ability of guanidine to selectively regulate transcription *in vitro* by measuring the extent of transcript elongation of several guanidine-I riboswitch constructs in the presence of guanidine or various analogs. Each construct has the potential to form a strong intrinsic terminator stem, but ligand binding presumably results in a stable aptamer structure that precludes formation of the terminator. In single-round transcription reactions with a riboswitch construct derived from the region upstream of

the *smr* gene from the bacterium *Desulfotomaculum ruminis* (**Figure 3-11A**), the inclusion of guanidine ranging from 1 to 100 mM yields progressively greater amounts of full-length RNA transcript (**Figure 3-11B**). James Nelson observed similar results for the guanidine-I riboswitches associated with the *Clostridium difficile emrE* (**Figure 3-12A**) and *B. subtilis ykkC* (**Figure 2-12B**) genes. Mutant construct M5, which carries a nucleotide change at a highly conserved position, exhibits a complete loss of guanidine responsiveness (**Figure 3-11C**). Likewise, mutations M6 and M7 predictably diminish and restore, respectively, the guanidine-responsive transcription of the *smr* construct.

Notably, guanidine has no effect on the extent of transcript elongation for a riboswitch that recognizes a different ligand (**Figure 3-12C**; Kim et al., 2015). Finally, the addition of structural analogs of guanidine, including urea, does not lead to transcript elongation (**Figure 3-11D**). These data demonstrate that the riboswitches tested are capable of specifically recognizing guanidine *in vitro* and that ligand binding by the RNA is sufficient to control the expression of downstream genes.

A urea carboxylase enzyme prefers guanidine over urea

We speculated that the genes controlled by guanidine-I riboswitches represent a novel super-regulon that is capable of reducing guanidine toxicity, either by its export or degradation. Therefore, Madeline Sherlock chose to look more closely at predicted urea carboxylase genes (Kanamori et al., 2004), which are frequently controlled by guanidine-I riboswitches (**Figure 3-1B**). Urea carboxylases and ureases have been proposed to catalyze the two major routes for the biological decomposition of urea. However, this redundancy seems unnecessary, and it has been previously hypothesized that one of these

protein families might have an alternative purpose (Hausinger, 2004). Intriguingly, urea carboxylase genes are commonly associated with guanidine-I riboswitches, whereas ureases never appear to be controlled by this riboswitch class. Therefore, we speculated that urea carboxylases might prefer guanidine as a substrate (**Figure 3-13A**) and initiate its degradation in a pathway that is similar to that currently proposed for urea.

The only previously characterized prokaryotic urea carboxylase, from the Proteobacterium *Oleomonas sagaranensis*, is encoded by a gene located immediately downstream of a mini-*ykkC* RNA. As noted above, guanidine-I riboswitches and the noncoding mini-*ykkC* and *ykkC*-III motif RNAs might have similar gene control functions. Madeline Sherlock was able to demonstrated that the *O. sagaranensis* urea carboxylase can indeed use both urea and guanidine as substrates (**Figure 3-13B**). Importantly, the K_M for urea is in the millimolar range (7.2 mM reported by Kanamori et al., 2004; 5.2 mM observed in this study, **Figure 3-13C**), whereas the K_M for guanidine improves to 0.19 mM. Thus, the enzyme carboxylates guanidine with a catalytic efficiency roughly 40-fold better than with urea. Since the maximum velocity and k_{cat} are only marginally higher compared to urea, this increased efficiency is presumably due to the enhanced affinity of the enzyme for guanidine.

Urea carboxylases are thought to have been introduced to eukaryotes by a horizontal gene transfer event from Proteobacteria (Strope et al., 2011). An x-ray structural model of the carboxylase domain of the urea amidolyase from the eukaryote *Kluyveromyces lactis* has been solved previously (Fan et al., 2012). Based on this structure, it is hypothesized that a protonated aspartic acid residue near the active site, which is conserved in the *O. sagaranensis* carboxylase, forms a hydrogen bond with the oxygen atom of urea. For this

to be true, however, the side-chain carboxylic acid of aspartate would have to undergo a pK_a shift of nearly 4 units. In contrast, Madeline Sherlock hypothesized that this residue could more easily form an ionic interaction or a hydrogen bond with guanidine without undergoing a shift in pK_a. This mechanism would explain the apparent enhanced affinity for guanidine over urea as described above. These observations suggest that diverse species of bacteria and eukarya express carboxylase enzymes that have evolved to degrade guanidine.

A riboswitch-controlled transport protein selectively binds guanidine

One of the most intriguing aspects of the guanidine-I riboswitch class is its association with genes predicted to code for small multidrug resistance (SMR) proteins. This family is subdivided into small multidrug proteins (SMP), suppressor of *groEL* mutation (SUG) proteins, and paired small multidrug resistance (PSMR) proteins. The latter group includes the proteins YkkC and YkkD in *B. subtilis* (Bay et al., 2008), which are regulated by a guanidine-I riboswitch (**Figures 3-3** and **3-10**). To date, most studies have been performed either on the SMP *E. coli* EmrE protein, which has been found to transport a wide variety of positively charged small molecules (Bay and Turner, 2012; Yerushalmi et al., 1995), or on the *E. coli* SugE protein, which has been suggested to transport considerably fewer compounds (Chung and Saier Jr., 2002).

To explore how SMR proteins encoded by genes controlled by guanidine-I riboswitches differ from the well-studied *E. coli* EmrE transporter, Randy Stockbridge constructed a phylogenetic tree of these proteins (**Figure 3-13D**). The riboswitch almost always controls a pair of SMR proteins, one of which appears to be EmrE-like and the

other SugE-like. Very frequently, their genes are found in a single operon, which suggests that heterodimers of EmrE-like and SugE-like proteins might form naturally. Additionally, when a guanidine-I riboswitch controls only a single copy of an *emrE*- or *sugE*-like gene, its putative partner is always found elsewhere in the genome, most often controlled by another guanidine riboswitch. In rare instances, guanidine-I riboswitches also control a single SMR gene that is highly similar to *E. coli sugE*. In contrast, genes that are highly similar to *E. coli emrE* are never controlled by the riboswitch.

We speculated that SMR proteins whose expression is controlled by guanidine-I riboswitches might naturally function as guanidine transporters. In a preliminary investigation of this hypothesis, Randy Stockbridge examined a representative SugE protein (called *Clo* SugE) that is derived from a *Clostridiales* bacterium (oral taxon 876) and controlled by a guanidine-I riboswitch. The distribution of positive charges on the loops suggested that this protein was likely to assemble as a homo-complex (Rapp et al., 2006), and size exclusion chromatography of the purified protein supported this inference. A dose-dependent increase in the intrinsic fluorescence of tryptophan residues within the *Clo* SugE protein was observed upon the addition of guanidine, demonstrating that it binds guanidine (**Figure 3-13E**). The K_D for this interaction is ~1 mM, as estimated by fitting the fluorescence data to a single-site binding model (**Figure 3-13F**). Importantly, a similar urea titration has no effect on *Clo* SugE fluorescence, and guanidine titration has no effect on an unrelated fluoride transporter with a similar number of tyrosine residues (data not shown).

Finally, Randy Stockbridge sought to determine whether *Clo* SugE was capable of selectively recognizing guanidine by monitoring the affinity of their interaction in the

presence of 1 mM tetraphenylphosphonium (TPP^+), a typical substrate of the *E. coli* EmrE protein. A concentration of TPP^+ at ~1,000-fold higher than its K_D for an *E. coli* EmrE (Morrison and Henzler-Wildman, 2014) fails to disrupt guanidine binding (**Figure 3-13G**). These results suggest that this riboswitch controlled SMR protein selectively recognizes guanidine and not the bulky, cationic substrates known to be transported by distal members of this general family of proteins.

Free guanidine is naturally produced by bacteria

The findings described above demonstrate the existence of riboswitches that selectively respond to free guanidine, and suggest that proteins are widely produced by bacteria that degrade and transport guanidine. However, there are no previous reports regarding the natural production of guanidine at biologically-relevant concentrations (Mitchell, 1985, 1987). It seems unlikely that an ancient riboswitch and its associated super-regulon would respond to an unnatural compound. Therefore, we speculated that guanidine might be produced through natural metabolic processes under certain conditions, and that its presence had not yet been detected, perhaps due to its small size or its perceived irrelevance. Accordingly, we sought to identify conditions where a guanidine riboswitch reporter is expressed even though no external guanidine is supplied.

If bacteria experience conditions under which guanidine is produced, the compound might be ejected from cells into the growth medium. Therefore, I chose an *E. coli* strain that lacks the *tolC* gene was chosen to examine our hypothesis because the TolC protein is a well-known exit channel for small toxic molecules in Gram-negative bacteria (Koronakis, 2003). Indeed the Δ*tolC* strain, but not the WT strain, exhibits robust WT

riboswitch reporter gene expression when grown overnight in minimal medium (**Figure 3-14A**). The M4 construct (**Figure 3-5A**), which carries a single mutation in the aptamer domain, exhibits lower induction by guanidine as expected. These data suggest that the riboswitch ligand is synthesized by *E. coli* under these conditions, and accumulates in cells that are deficient in exporting small toxic compounds.

James Nelson analyzed the contents of these bacteria via high-resolution liquid chromatography-mass spectrometry (LC-MS) and observed the presence of a compound with a mass-to-charge ratio equivalent to that of guanidine (**Figure 3-14B**). This compound also had a similar retention time to that of guanidine (**Figure 3-14C**). James Nelson speculated that the modest difference in these retention times might be due to the intrinsic differences of a pure guanidine solution and a complex extract sample. To further confirm that this compound was, in fact, guanidine, he treated cell extracts with di-tert-butyl dicarbonate (Boc-anhydride) to convert guanidine into Boc-guanidine. A new compound with the identical mass-to-charge ratio and similar retention time as authentic Boc-guanidine was observed. Upon mixing the Boc-anhydride-treated extract and Boc-guanidine, a single LC-MS peak was observed (**Figure 3-14C**), indicating that these peaks both correspond to Boc-guanidine. These results demonstrate that *E. coli* is capable of endogenously producing guanidine.

Discussion

It has been proposed that the cognate ligands for widespread orphan riboswitch classes are likely to be involved in underappreciated, but fundamental, aspects of biology (Breaker, 2011). The data described in the current report supports the conclusion that free guanidine is the natural ligand for the predominant subtype of *ykkC* motif RNAs, the most common and longest unresolved orphan riboswitch candidate. Despite being generated over 150 years ago, only in rare instances has guanidine been considered to be a biologically relevant compound. However, my and James Nelson's results (**Figure 3-13**) demonstrate that guanidine is naturally produced by bacteria. Perhaps many bacteria generate free guanidine as they utilize suboptimal energy sources when preferred sources such as glucose are in short supply.

Guanidine-I riboswitches, and the structurally unrelated mini-*ykkC* and *ykkC*-III motif RNAs (Weinberg et al., 2010) that function as guanidine riboswitches (Sherlock and Breaker, 2017; Sherlock et al., 2017), control a super-regulon that includes genes for proteins that appear to detoxify free guanidine by degradation or export. Given the widespread distribution of guanidine riboswitches in bacteria, free guanidine might have been biologically relevant in early life forms. Notably, guanidine is one of a number of compounds that could possibly have been part of a prebiotic pathway for purine synthesis (Becker et al., 2016). Guanidine degradation in bacteria could occur in the same manner proposed for urea, wherein previously annotated urea carboxylase enzymes preferentially use guanidine as a substrate (**Figure 3-13A**). Genes coding for such 'guanidine carboxylases' are almost always found associated with two families of uncharacterized proteins. It seems likely that these other proteins play some additional role in guanidine

metabolism. Notably, the product of carboxylation of guanidine, carboxyguanidine, is a close analog of allophanate. Therefore, Madeline Sherlock hypothesized that proteins annotated as allophanate hydrolases whose expression levels are controlled by guanidine riboswitches presumably use carboxyguanidine as a substrate for subsequent decomposition.

Another class of genes commonly regulated by these riboswitches encode arginases, which typically catalyze the conversion of arginine into ornithine. Examination of the guanidine riboswitches that are upstream of the genes encoding these arginases reveals that they most likely activate gene expression in the presence of guanidine. When controlled by the riboswitch, these genes are always found in an operon with the genes *hypA* and *hypB*, which have been implicated in the insertion of Ni^{2+} into hydrogenases in *E. coli* (Lutz et al., 1991). Since arginases can also utilize a Ni^{2+} ion at their active site (Viator et al., 2008), we speculate that *hypA* and *hypB* may be involved in the insertion of Ni^{2+} or some other cation into these enzymes. I hypothesized that guanidine-induced expression of this operon helps cells overcome inhibition of arginase activity by free guanidine. Alternatively, these enzymes could conceivably catalyze the difficult reverse reaction, where ornithine or 5-hydroxynorvaline is directly converted to arginine via the addition of guanidine.

Guanidine riboswitches also commonly control the expression of numerous transporters. Many of these are predicted to import nitrate, sulfate, or bicarbonate. It seems possible that these proteins might transport bicarbonate to facilitate the carboxylation of guanidine and its subsequent metabolism. Indeed, genes encoding such transporters are often found in an operon with guanidine carboxylases genes. The

majority of the remaining transporters whose expression levels are controlled by guanidine riboswitches are predicted SMR transporters. Certain members of this protein superfamily, such as the *E. coli* proteins EmrE and SugE have been extensively studied as model transporters due to their small size. EmrE in particular has been demonstrated to transport a wide variety of positively charged small molecules ranging from TPP^+ to glycine betaine (Bay and Turner, 2012; Yerushalmi et al., 1995). Randy Stockbridge demonstrated that one of the SMR proteins whose expression is controlled by a guanidine riboswitch selectively recognizes guanidine. Her data further suggest that SMR transporters whose expression levels are controlled by guanidine-I riboswitches usually assemble as heterodimers of SugE- and Emr-E like subunits and that their biological role is to function as guanidine exporters. This hypothesis is supported by my observation that *B. subtilis* cells wherein this transporter is knocked out have a greater sensitivity to guanidine (**Figure 3-3D**) and a higher cellular concentration of the riboswitch ligand (**Figure 3-3E and 3-3F**).

In *E. coli*, the SugE protein is controlled by the mini-*ykkC* motif, suggesting it might also transport guanidine. Interestingly, the *sugE* gene was first identified as a suppressor of a *groEL* mutant gene, a gene that encodes one of a family of molecular chaperones that assists in the proper folding of proteins (Greener et al., 1993). An intriguing possibility is that SugE reduces the cellular concentration of guanidine, a known protein destabilization agent, thereby reducing the need for GroEL to chaperone protein folding.

It is also important to note that *emrE* genes not controlled by guanidine-I riboswitches are commonly associated with horizontal gene transfer elements and are relatively recently diverged (Bay et al., 2008). In contrast, genes that encode SMR proteins that are

controlled by these RNAs are only rarely found within 10 kb of an integrase or transposase. These observations suggest that a mutation may have occurred in an ancestral guanidine transporter that allowed it to transport a wider variety of cationic compounds, leading to its rapid transfer amongst bacteria and subsequent contribution to antibiotic resistance.

As noted earlier, Madeline Sherlock discovered that approximately 25% of *ykkC* motif RNAs conform to subtype 2 (**Figure 3-2A**). Members of this subtype are commonly associated with genes necessary for *de novo* purine metabolism and branched-chain amino acid biosynthesis. These RNAs are distinguished by nucleotide variations at the sites where the structure of guanidine-I riboswitches modulates in the presence of guanidine. Therefore, subtype 2 *ykkC* motif RNAs appear to have adapted to recognize one or more unknown ligands that relate to the metabolic processes relevant to their associated genes. Experiments by Madeline Sherlock and collogues have now shown that subtypes 2A, 2B, and 2C of the *ykkC* motif RNAs selectively recognize ppGpp, PRPP, and (d)ATP, respectively (Sherlock et al., 2018a; Sherlock et al., 2018b; Sherlock et al., 2018c).

Material and Methods

Chemicals and reagents. Guanidine was used in its hydrochloride salt form in all experiments and was purchased from Sigma-Aldrich. Other fine chemicals were also purchased from Sigma-Aldrich with the exceptions of hydroxyguanidine hemisulfate (Santa Cruz Biotech) and guanidinosuccinic acid (Alfa Chemistry). [γ-^{32}P]- and [α-^{32}P]-ATP were purchased from Perkin Elmer and used within two weeks of receipt. Bulk chemicals were purchased from J.T. Baker and enzymes from New England Biolabs, unless otherwise noted. All solutions were prepared using deionized water (dH_2O) and were either autoclaved or filter sterilized (using 0.22 μm filters, Millipore) prior to use. DNA oligonucleotides were purchased from Sigma-Aldrich and Integrated DNA Technologies. A list of oligonucleotides used in this study can be found in **Table 3-1**.

Bioinformatics. Additional homologs of *ykkC* motif RNAs were identified by using Infernal 1.1 (Nawrocki and Eddy, 2013) based on the previously published *ykkC-yxkD* motif alignment (Barrick et al., 2004). Iterative searches and re-alignment of a library of genomic sequences obtained from bacterial sequences deposited in RefSeq version 63 (Pruitt et al., 2007) and environmental sequences from the Human Microbiome Project (Gevers, 2012), IMG/M (Markowitz et al., 2008), MG-RAST (Meyer, 2008), CAMERA (Sun, 2011), and GenBank (Benson et al., 2008) resulted in the identification of a group of nearly 2000 hits. Genes were predicted to be controlled by the adjacent RNA motif if they were downstream and within 700 nucleotides, or if they were predicted to be in an operon controlled by the RNA motif. These genes were then sorted based on predicted function into various classes as presented in **Figure 3-1B** and **Figure 3-2B**.

SMR protein phylogenies were constructed from 221 bacterial sequences, including 182 riboswitch-controlled SMRs and ~39 SMRs that have been experimentally implicated in drug resistance (Bay et al., 2008; Cruz et al., 2013; He et al., 2011; Kovacevic et al., 2016; Nasie et al., 2012). Sequences were aligned using MUSCLE (Edgar, 2004), and variable loop regions were trimmed from the alignment using JalView. The unrooted phylogeny was constructed using PhyML 3.0 (Guindon et al., 2010).

Reporter assays and knockout construction. Initially, a riboswitch reporter construct consisting of the native *B. subtilis ykkC* gene promoter and riboswitch fused upstream of the *E. coli lacZ* gene (**Figure 3-3**) was created by using methods as described previously (Sudarsan et al., 2003) and integrated into the *amyE* locus of *B. subtilis* (Meyer et al., 2011). The construct with the native promoter was used for growth condition screening (**Figure 3-4**), but the promoter was subsequently replaced with the *B. subtilis lysC* promoter via PCR, restriction digest, and subsequent cloning into the vector pDG1661. This latter construct, which avoids any possible complications from *ykkC* gene natural promoter regulation, was used for all other *B. subtilis* experiments. Incorporation of various mutations was carried out via PCR followed by restriction digestion and cloning. The *K. pneumoniae* riboswitch reporter construct (**Figure 3-5A**) was designed to include the *lysC* promoter from *E. coli* and the entire 5' untranslated region of the *tauA* gene fused directly to the start codon for the full coding region of the *lacZ* reporter gene. The resulting vector pRS414 construct was transformed into *E. coli* BL21DE3 cells. Screening of various conditions via a phenotype microarray was performed by Biolog,

Inc. using proprietary solutions and media. Follow-up reporter experiments were performed by inoculating either *B. subtilis* strain 168 1A1 (Δ*trp*) or *E. coli* BW25113 transformed with the riboswitch reporter construct into Lysogeny Broth (LB) or minimal medium (GMM for *B. subtilis* or M9 for *E. coli*) supplemented with various concentrations of guanidine or other compounds as indicated. Bacteria were grown overnight at 37ºC, with shaking if using the *lysC* promoter, or without if using the native promoter. Synthetic DNAs used in cloning are listed in **Table 3-1**.

Preparation of radiolabeled RNA constructs for in-line probing. Templates for RNA transcription were assembled via extension of overlapping synthetic DNAs (**Table 3-1**) using SuperScript II reverse transcriptase (Thermo Fisher Scientific). Each RNA construct was generated by transcription for four hours at 37ºC in a 100 µL reaction containing 80 mM HEPES-KOH (pH 7.5 at 23ºC), 24 mM $MgCl_2$, 2 mM spermidine, 40 mM DTT, and T7 RNA polymerase (16 U μL^{-1}). Each desired RNA product was purified via denaturing PAGE, the gel slice containing the full-length RNA was excised, and the RNA extracted via crush-soak treatment (200 mM NaCl, 10 mM Tris [pH 7.5 at 23ºC], 1 mM EDTA; 400 µL) for 30 minutes at room temperature. The purified RNAs were subsequently precipitated by addition of 1 mL ethanol, pelleted via centrifugation, dried via rotary evaporation (DNA SpeedVac), and resuspended in 20 µL dH_2O. RNAs were dephosphorylated via rAPid alkaline phosphatase (Roche) treatment according to the manufacturer's instructions. 8 pmol of RNA was radiolabeled using T4 polynucleotide kinase in a 20 µL reaction volume containing 25 mM CHES [pH 9.0 at 23ºC], 5 mM

$MgCl_2$, 3 mM DTT, and 20 μCi [γ-^{32}P]-ATP. The resulting radiolabeled RNAs were purified again as described above before being subjected to in-line probing.

Single-round transcription termination assays. DNA constructs for single-round *in vitro* transcription were designed to carry the riboswitch of interest spanning from the predicted natural transcription start site to 40 (*C. difficile* and *D. ruminis* constructs) or 50 (*B. subtilis* construct) nucleotides following the terminator stem. The promoter from the *pfl* gene in *Clostridium beijerinckii* NCIMB 8052 was used due to its previously demonstrated function in similar assays (Kim et al., 2015). A mutation was incorporated in the region of the *B. subtilis* construct before the *ykkC* aptamer so that it contained no cytidines, while the *C. difficile* and *D. ruminis* constructs were naturally cytidine-deficient. Finally, 36 additional nucleotides were added to the 3′ termini of the *C. difficile and D. ruminis* constructs to aid in both resolution of the full-length and terminated products and construction of transcription templates. Synthetic DNA templates based on these design features were then constructed by extension of overlapping, synthetic single-stranded oligonucleotides (Integrated DNA Technologies) followed by PCR amplification of the final product. Approximately 2 pmol of the resulting, purified DNA template was added to a transcription initiation mixture (20 mM Tris-HCl [pH 8.0 at 23ºC], 20 mM NaCl, 7 mM $MgCl_2$, 0.1 mM EDTA, 10 μg mL^{-1} bovine serum albumin [BSA], 130 μM ApA dinucleotide, 1% glycerol, 0.04 U μL^{-1} *E. coli* RNA polymerase holoenzyme, 2.5 μM GTP, 2.5 μM UTP, and 1.5 μM ATP). Approximately 8 μCi [α-^{32}P]-ATP was added to the 90 μL transcription reaction and transcription was allowed to proceed at 37ºC for 10 minutes, leading to formation of a stalled polymerase complex at

the first cytidine nucleotide of each transcript. The reaction mixture was then distributed in 8 μL aliquots into separate tubes to which 1 μL of a 10x solution of the ligand of interest was added. The resulting solutions were gently mixed and centrifuged before being placed on ice. Subsequently, 1 μL of 10x elongation buffer (20 mM Tris-HCl [pH 8.0 at 23ºC], 20 mM NaCl, 7 mM $MgCl_2$, 0.1 mM EDTA, 1 mg mL^{-1} heparin) was added. The 10x elongation buffer for the *C. difficile* and *D. ruminis* constructs also contained 1.5 mM each of ATP, GTP, and CTP, as well as 500 μM UTP. The 10x elongation buffer for the *B. subtilis* construct contained 3 mM each of ATP, GTP, and CTP, and 1 mM UTP. The percentages of elongated RNA transcripts were calculated as previously described (Kim et al., 2007). Experimental conditions for single-round transcription of the ZTP riboswitch from *C. beijerinckii* have been described previously (Kim et al., 2015).

Preparation and analysis of an annotated urea carboxylase. The full-length urea carboxylase (*uca*) gene from *O. sagaranensis* was cloned into the multiple cloning site 1 of the pETDuet-1 vector (Novagen) using the BamHI and HindIII restriction sites. The recombinant protein was designed such that it contained a His_6 tag at its N terminus. The resulting plasmid was then transformed into *E. coli* BL21(DE3) and cultured in Terrific Broth (IBI Scientific) supplemented with 100 mg L^{-1} ampicillin. Once cells reached an OD_{600} of roughly 0.2, 15 mg L^{-1} D-biotin was added. At an OD_{600} of between 0.8 and 1.0, expression of the *uca* gene was induced by the addition of 0.5 mM isopropyl-β-D-thiogalactopyranoside (IPTG) and incubated at 16ºC for 18 h. Cells were harvested by centrifugation and lysed using a microfluidizer (Microfluidics) at 15,000 psi. The lysate was then clarified by ultracentrifugation at 40,000 RCF for 40 min at 4ºC. His_6-tagged

protein was purified first via nickel-affinity chromatography using nickel nitrolotriacetic acid (Ni-NTA) agarose (Anatrace) and then by strong anion exchange chromatography (HiTrap Q HP, GE Healthcare), both performed on an AKTA purification system. The purified protein was subsequently exchanged into buffer containing 25 mM Tris-HCl (pH 8.0 at 23ºC), 100 mM NaCl, and 0.1 mM TCEP (tris-(2-carboxyethyl)phosphine). The resulting protein sample was concentrated via dialysis, flash-frozen in liquid nitrogen, and stored at -80ºC.

Catalytic activity of the *O. sagaranensis* carboxylase enzyme was determined by coupling the carboxylation reaction with a lactate dehydrogenase and pyruvate kinase system as described previously (Kanamori et al., 2004). Briefly, conversion of ATP to ADP by carboxylase is coupled to conversion of NADH to NAD^+ by lactate dehydrogenase and pyruvate kinase, which can be followed by measuring the absorbance of NADH at 340 nm. Guanidine or urea was added at the indicated concentration to the reaction mixture [50 mM HEPES (pH 7.5 at 23ºC), 50 mM KCl, 8 mM $MgCl_2$, 0.5 mM phosphoenolpyruvate, 8 mM $NaHCO_3$, 0.4 mM NADH, 5 U of pyruvate kinase (Sigma-Aldrich), and 1 U of lactate dehydrogenase (Sigma-Aldrich)] to yield a final reaction volume of 50 μL. The reactions were allowed to incubate at room temperature (~23ºC) for 45 minutes. ATP was then added to a final concentration of 1 mM to reactions incubated at 30ºC, and absorbance measurements at 340 nm immediately commenced using a Biotek Synergy H1 plate reader. Three technical replicates were performed per trial for each concentration of substrate. Absorbance as a function of time was recorded and the slope of each data set was used to determine the rate of substrate carboxylation. Velocity is reported such that 1 U of activity is defined as 1 μmol NAD^+ formed per

minute. The velocity for the no-substrate control (0.5 U mg^{-1}) was subtracted prior to determination of kinetic parameters. V_{max}, K_M, and k_{cat} were determined by fitting a plot of velocity vs. substrate concentration using GraphPad Prism 6 assuming Michaelis-Menten kinetics.

Expression and purification of SugE. *E. coli* transformed with the *sugE*-bearing plasmid were induced with 1 mM IPTG at an OD_{600} of 1.5, and protein expression was allowed to proceed for 3 hr at 37°C. Cells were lysed by sonication and the resulting mixture was extracted with 60 mM n-decyl-β-D-maltopyranoside (DM) for 2 hr. Cell debris was pelleted, the supernatant was passed over a cobalt affinity resin (Talon, 1 mL per L of culture), and the resin was subsequently washed first with a mixture of 100 mM NaCl, 20 mM Tris (pH 7.5 at 23ºC), and 5 mM DM (column buffer, CB), then CB supplemented with 20 mM imidazole, and finally eluted with CB supplemented with 400 mM imidazole. Histidine tags were removed by 1 hr treatment with endoproteinase Lys-C (Roche, 0.2 U mg^{-1} protein). The desired protein was further purified using a Superdex200 gel filtration column with an aqueous solution containing 25 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) (pH 7.5 at 23ºC), 150 mM NaCl, and 5 mM DM.

Tryptophan fluorescence assay. Protein samples were titrated with guanidinium chloride dissolved in 150 mM NaCl and 25 mM HEPES (pH 7.5 at 23ºC), with 5 mM DM, and incubated for 2 min to allow the sample to reach an equilibrium value before excitation at 280 nm. Emission was monitored from 200-400 nm, and titration curves

were constructed from the emission spectrum peak (333 nm), which typically increased by ~50% over the course of the experiment. For experiments comparing binding in the presence and absence of 1 mM $TPP^{+,}$ the excitation wavelength was 293 nm. Each titration experiment was repeated in triplicate, and data were fit to a single-site binding model using SigmaPlot. For comparison between experiments, data were normalized to the maximum fluorescence value obtained from the regression analysis.

LC-MS analysis of bacterial cell extracts. Bacteria were harvested via centrifugation, the supernatant removed, and the pellet dried via lyophilization. The bacteria were lysed via the addition of 0.1 culture volumes of 50% aqueous methanol (v/v). Insoluble cell debris or salts were pelleted via centrifugation and the supernatant removed, lyophilized to dryness, and resuspended in 0.01 culture volumes of dH_2O.

Extract samples were analyzed via LC-MS first by injection onto a Hypercarb porous graphitic carbon column (2.1 mm x 100 mm, particle size 3 μm, Thermo Fisher Scientific). The column was pre-equilibrated with 2% acetonitrile (ACN) for five minutes prior to injection, followed by elution with 2% ACN over 15 minutes, gradually increasing to 99% ACN over a further 20 minutes. The mass-to-charge ratio of compounds within the eluent was determined using an Agilent 6550 qToF mass spectrometer. Data were analyzed using the Agilent Mass Hunter software. Peaks were considered to have the same mass-to-charge ratio as guanidine if they were within 10 ppm of the calculated ratio. To confirm that this peak corresponded to guanidine, cell extracts or authentic guanidine alone were treated with 10 mM t-butoxycarbonyl anhydride (Boc-anhydride) in the presence of 100 mM NaOH for three hours at room

temperature to convert any guanidine into Boc-guanidine. The resulting reaction mixtures were lyophilized, extracted with 50% aqueous methanol (v/v), resuspended in dH_2O, and analyzed via LC-MS as described above. Peaks corresponding to the expected mass of Boc-guanidine were observed in both samples, but they had slightly different LC retention times. These peaks coalesced into a single peak upon mixing the two samples, indicating that the differences in retention times are due to sample effects from the separate preparations.

References

Baker JL, Sudarsan N, Weinberg Z, Roth A, Stockbridge RB, Breaker RR. 2012. Widespread genetic switches and toxicity resistance proteins for fluoride. *Science* **335:** 233-235.

Barrick JE, Corbino KA, Winkler WC, 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. USA* **101:** 6421-6426.

Bay DC, Rommens KL, Turner RJ. 2008. Small multidrug resistance proteins: a multidrug transporter family that continues to grow. *Biochim. Biophys. Acta* **1778:** 1814-1838.

Bay DC, Turner RJ. 2012. Small multidrug resistance protein EmrE reduces host pH and osmotic tolerance to metabolic quaternary cation osmoprotectants. *J. Bacteriol.* **194:** 5941-5948.

Becker S, Thoma I, Deutsch A, Gehrke T, Mayer P, Zipse H, Carell T. 2016. A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352:** 833-836.

Benson DA, Karsch-Mizrachi I, Lipman DJ, Ostell J, Wheeler DL. 2008. Genbank. *Nucleic Acids Res.* **36:** D25-D30.

Bierema S. 1909. Die assimilation von ammon-, nitrat- und amidstickstoff durch mikroorganismen. *Centr. Bakt. Par.* **2:** 672-726.

Block KF, Hammond MC, Breaker RR. 2010. Evidence for widespread gene control function by the *ydaO* riboswitch candidate. *J. Bacteriol.* **192:** 3983-3989.

Breaker RR. 2011. Prospects for riboswitch discovery and analysis. *Mol. Cell* **43:** 867-879.

Chung YJ, Saier Jr. MH. 2002. Overexpression of the *Escherichia coli sugE* gene confers resistance to a narrow range of quaternary ammonium compounds. *J. Bacteriol.* **184:** 2543-2545.

Cruz A, Micaelo N, Felix V, Song JY, Kitamura S, Suzuki S, Mendo S. 2013. *sugE*: a gene involved in tributyltin (TBT) resistance of *Aeromonas molluscorum* Av27. J. Gen. Appl. Microbiol. **59:** 39-47.

Dambach M, Sandoval M, Updegrove TB, Anantharaman V, Aravind L, Waters LS, Storz G. 2015. The ubiquitous *yybP-ykoY* riboswitch is a manganese-responsive regulatory element. *Mol. Cell* **57:** 1099-1109.

Dann CE, 3rd Wakeman CA, Sieling CL, Baker SC, Irnov I, Winkler WC. 2007. Structure and mechanism of a metal-sensing regulatory RNA. *Cell* **130:** 878-892.

Ebisuno T, Takimoto M. 1981. Examination on the biodegradability of guanyl compounds. *Eisei Kagaku* **27:** 156-162.

Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res*. **32:** 1792-1797.

Fan C, Chou CY, Tong L, Xiang S. 2012. Crystal structure of urea carboxylase provides insights into the carboxyltransfer reaction. *J. Biol. Chem*. **287:** 9389-9398.

Frecker NL. 2013. Regulation of the putative *ykk*CD riboswitch by tetracycline and related antibiotics in *Bacillus subtilis*. Master's Dissertation, Ball State University.

Gevers D. 2012. The human microbiome project: a community resource for the healthy human microbiome. *PLoS Biol*. **10:** e1001377.

Greene RF, Pace CN. 1974. Urea and guanidine hydrocholoride denaturation of ribonuclease, lysozyme, α-chymotrypsin, and β-lactoglobulin. *J. Biol. Chem*. **249:** 5388-5393.

Greener T, Govezensky D, Zamir A. 1993. A novel multicopy supressor of a groEL mutation includes two nested open reading frames transcribed from different promoters. *EMBO J*. **12:** 889-896.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordjik W, Gascuel O. 2010. New algorithims and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst. Biol*. **59:** 307-321.

Hausinger RP. 2004. Metabolic versatility of prokaryotes for urea decomposition. *J. Bacteriol*. **186:** 2520-2522.

Gusarov I, Nudler E. 1999. The mechanism of intrinsic transcription termination. *Mol. Cell* **3:** 495-504.

He GX, Zhang C, Crow RR, Thorpe C, Chen H, Kumar S, Tsuchiya T, Varela MF. 2011. SugE, a new member of the SMR family of transporters, contributes to antimicrobial resistance in *Enterobacter cloacae*. *Antimicrob. Agents Chemother.* **55:** 3954-3957.

Hermann T, Westhof E. 1999. Docking of cationic antibiotics to negatively charged pockets in RNA folds. *J. Med. Chem*. **42:** 1250-1261.

Iwanoff NN, Awetissowa AN. 1931. Über die fermentative umwandlung des guanidins in harnstoff. *Biochem. Z*. **231:** 67-78.

Jack DL, Storms ML, Tchieu JH, Paulsen IT, Saier Jr. MH. 2000. A broad-specificity multidrug efflux pump requiring a pair of homologous SMR-type proteins. *J. Bacteriol*. **182:** 2311-2313.

Kanamori T, Kanou N, Atomi H, Imanaka T. 2004. Enzymatic characterization of a prokaryotic urea carboxylase. *J. Bacteriol*. **186:** 2532-2539.

Kihara H, Prescott JM, Snell EE. 1955. The bacterial cleavage of canavanine to homoserine and guanidine. *J. Biol. Chem*. **217:** 497-503.

Kim JN, Roth A, Breaker RR. 2007. Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2'-deoxyguanosine. *Proc. Natl. Acad. Sci. USA* **104:** 16092-16097.

Kim PB, Nelson JW, Breaker RR. 2015. An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol. Cell* **57:** 317-328.

Koronakis V. 2003. TolC – the bacterial exit duct for proteins and drugs. *FEBS Lett*. **555:** 66-71.

Kovacevic J, Ziegler J, Walecka-Zacharska E, Reimer A, Kitts DD, Gilmour MW. 2016. Tolerance of *Listeria monocytogenes* to quaternary ammonium sanitizers is mediated by a novel efflux pump encoded by *emrE*. *Appl. Environ. Microbiol*. **82:** 939-953.

Li S, Hwang XY, Stav S, Breaker RR. 2016. The *yjdF* riboswitch candidate regulates gene expression by binding diverse azaaromatic compounds. *RNA* **22:** 530-541.

Lutz S, Jacobi A, Schlensog V, Böhm R, Sawers G, Böck A. 1991. Molecular characterization of an operon (*hyp*) necessary for the activity of the three hydrogenase isoenzymes in *Escherichia coli*. *Mol. Microbiol*. **5:** 123-135.

Mandal M, Lee M, Barrick JE, Weinberg Z, Emilsson GM, Ruzzo WL, Breaker RR. 2004. A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306:** 275-279.

Marchais A, Duperrier S, Durand S, Gautheret D, Stragier P. 2011. CsfG, a sporulation-specific, small non-coding RNA highly conserved in endospore formers. *RNA Biol*. **8:** 358-364.

Marescau B, Nagels G, Possemiers I, De Broe ME, Becaus I, Billiouw JM, Lornov W, De Deyn PP. 1997. Guanidino compounds in serum and urine of nondialyzed patients with chronic renal insufficiency. *Metabolism* **46:** 1024-1031.

Markowitz VM, Ivanova NN, Szeto E, Palaniappan K, Chu K, Dalevi D, Chen IM, Grechkin Y, Dubchak I, Anderson I, et al. 2008. IMG/M: a data management and analysis system for metagenomes. *Nucleic Acids Res*. **36:** D534-D538.

Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N, Breaker RR. 2011. Challenges of ligand identification for riboswitch candidates. *RNA Biol*. **8:** 5-10.

Mitchell WR. 1985. Biodegradation of guanidinium by aquatic microorganisms. US Army Med. Res. and Dev. *Command Tech. Rep*. **8506:** 1-26.

Meyer MM, Hammond MC, Salinas Y, Roth A, Sudarsan N, Breaker RR. 2011. Challenges of ligand identification for riboswitch candidates. *RNA Biol*. **8:** 5-10.

Meyer F. 2008. The metagenomics RAST server - a public resource for the automatic phylogenetic and functional analysis of metagenomes. *BMC Bioinfo*. **9:** 386-393.

Mitchell WR. 1987. Biodegradation of guanidinium ion in aerobic soil samples. *Bull. Environ. Contam. Toxicol.* **39:** 974-981.

Morrison EA, Henzler-Wildman KA. 2014. Transported substrate determines exchange rate in the multidrug resistance transporter EmrE. J. *Biol. Chem*. **289:** 6825-6836.

Nasie I, Steiner-Mordoch S, Schuldiner S. 2012. New substrates on the block: clinically relevant resistances for EmrE and homologues. *J. Bacteriol*. **194:** 6766-6770.

Natelson S, Sherwin JE. 1979. Proposed mechanism for urea nitrogen reutilization: relationship between urea and proposed guanidine cycles. *Clin. Chem*. **25:** 1343.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29:** 2933-2935.

Nelson JW, Plummer MS, Blount KF, Ames TD, Breaker RR. 2015a. Small molecule fluoride toxicity agonists. *Chem. Biol*. **22:** 527-534.

Nelson JW, Sudarsan N, Phillips GE, Stav S, Lünse CE, McCown PJ, Breaker RR. 2015b. Control of bacterial exoelectrogenesis by c-AMP-GMP. *Proc. Natl. Acad. Sci. USA* **112:** 5389-5394.

Nelson JW, Sudarsan N, Furukawa K, Weinberg Z, Wang JX, Breaker RR. 2013. Riboswitches in eubacteria sense the second messenger c-di-AMP. *Nat. Chem. Biol*. **9:** 834-839.

Price IR, Gaballa R, Ding F, Helmann JD, Ke A. 2015. Mn^{2+}-sensing mechanisms of *yybP-ykoY* orphan riboswitches. *Mol. Cell* **57:** 1110-1123.

Pruitt KD, Tatusova T, Maglott DR. 2007. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts, and proteins. *Nucleic Acids Res*. **35:** D61-D65.

Pyle AM. 2002. Metal ions in the structure and function of RNA. *J. Biol. Inorg. Chem*. **7:** 679-690.

Rapp M, Granseth E, Seppala S, von Heijne G. 2006. Identification and evolution of dual-topology membrane proteins. *Nat. Struct. Mol. Biol*. **13:**112-116.

Regulski EE, Breaker RR. 2008. In-Line Probing Analysis of Riboswitches. *Methods Mol. Biol*. **419:** 53-67.

Reiss CW, Xiong Y, Strobel SA. 2017. Structural basis for ligand binding to the guanidine-I riboswitch. *Structure **25*:** 195−202.

Roth A, Winkler WC, Regulski EE, Lee BW, Lim J, Jona I, Barrick JE, Ritwik A, Kim JN, Welz R*, et al.* 2007. A riboswitch selective for the queuosine precursor $preQ_1$ contains an unusually small aptamer domain. *Nat. Struct. Mol. Biol*. **14:** 308-317.

Schuldiner SS. 2009. EmrE, a model for studying evolution and mechanism of ion-coupled transporters. *Biochim. Biophys. Acta* **1794:** 748-762.

Schuldiner SS. 2012. Undecided membrane proteins insert in random topologies. Up, down and sideways: it does not really matter. *Trends Biochem. Sci*. **37:** 215-219.

Sekowska A, Danchin A, Risler JL. 2000. Phylogeny of related functions: the case of polyamine biosynthetic enzymes. *Microbiology* **146:** 1815-1828.

Serganov A, Nudler, E. 2013. A decade of riboswitches. Cell *152*, 17-24.

Soukup GA, Breaker RR. (1999). Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5:** 1308-1325.

Strecker A. 1861. Untersuchungen über die chemischen Beziehungen zwischen Guanin, Xanthin, Theobromin, Caffeïn und Kreatinin. *Justus Liebigs Ann. Chem*. **118:** 151-177.

Strope PK, Nickerson KW, Harris SD, Moriyama EN. 2011. Molecular evolution of urea amidolyase and urea carboxylase in fungi. *BMC Evol. Biol*. **2011:** 11:80

Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. 2003. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev.* **17:** 2688-2697.

Sun S. 2011. Community cyberinfrastructure for advanced microbial ecology research and analysis: the CAMERA resource. *Nucleic Acids Res*. **39:** D546-D551.

Viator RJ, Rest RF, Hildebrandt E, McGee DJ. 2008. Characterization of *Bacillus anthracis* arginase: effects of pH, temperature, and cell viability on metal preference. *BMC Biochem*. **9:** 15.

Walker JB, Walker MS. 1959. The enzymatic reduction of hydroxyguanidine. *J. Biol. Chem.* **234:** 1481-1484.

Weinberg Z, Barrick JE, Yao Z, Roth A, Kim JN, Gore J, Wang JX, Lee ER, Block KF, 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.

Weinberg Z, Breaker RR. 2011. R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinfo*. **12:** 3.

Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR. 2010. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol*. **11:** R31.

Winkler WC, Nahvi A, Roth A, Collins JA, Breaker RR. 2004. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428:** 281-286.

Yarnell WS, Roberts JW. 1999. Mechanism of intrisnic transcription termination and antitermination. *Science* **284:** 611-615.

Yerushalmi H, Lebendiker M, Schuldiner S. 1995. EmrE, an *Escherichia coli* 12-kDa multidrug transporter, exchanges toxic cations, and H+ and is soluble in organic solvents. *J. Biol. Chem*. **270:** 6856-6863.

Figures and Tables

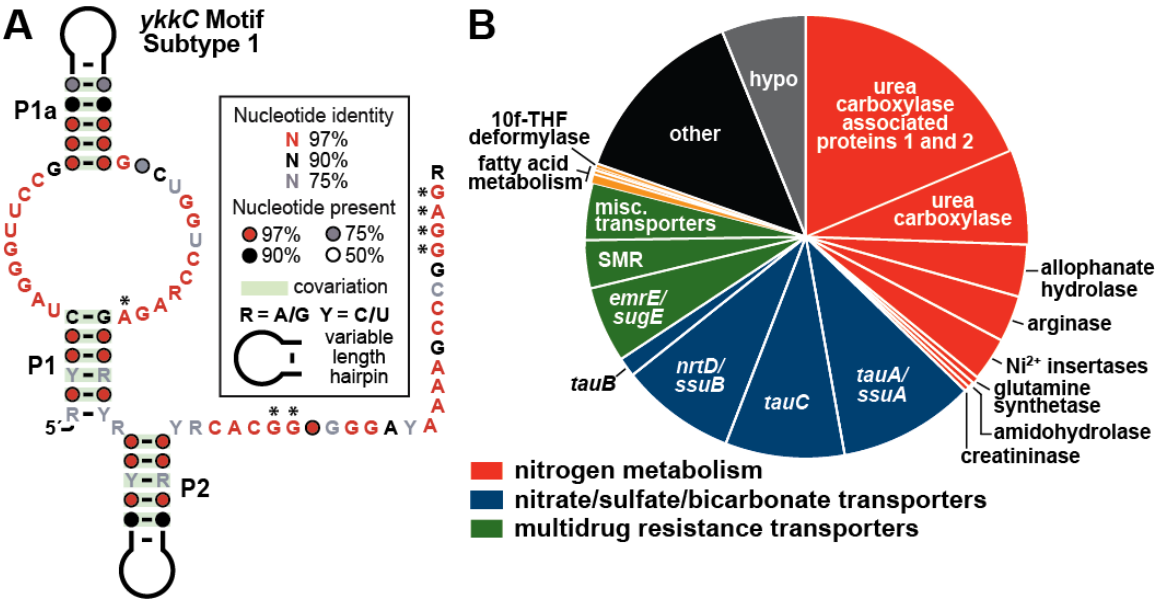

Figure 3-1. The *ykkC* motif is widespread in bacteria. (A) Consensus sequence and secondary structure model derived from ~1500 subtype 1 *ykkC* motif RNAs. Asterisks identify nucleotides that differ between subtypes 1 and 2. (B) Genes predicted to be controlled by subtype 1 *ykkC* motif RNAs.

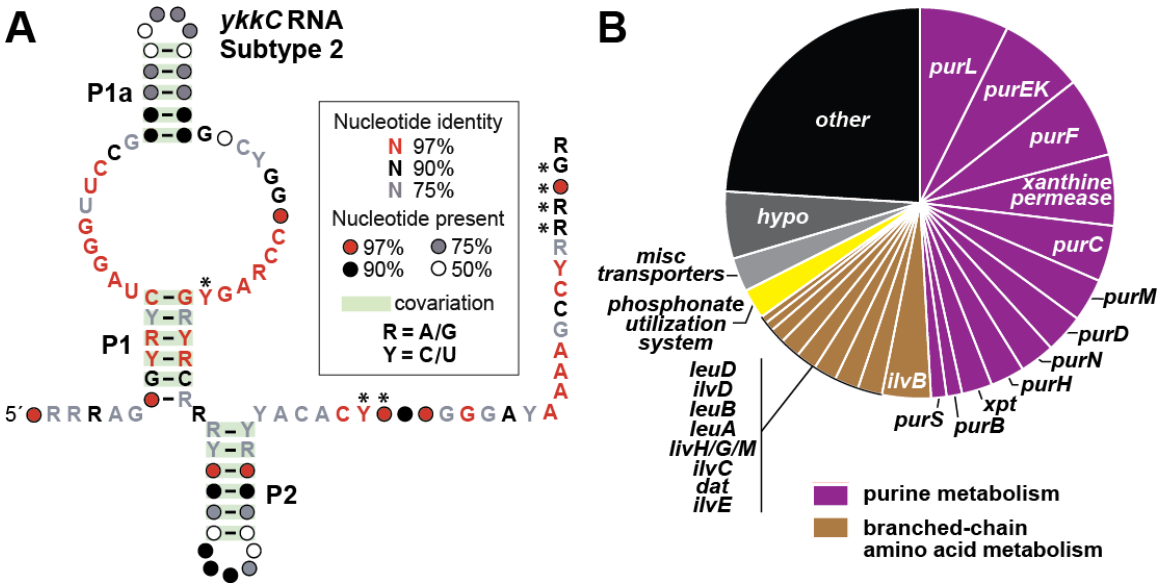

Figure 3-2. Subtype 2 *ykkC* motif RNAs associate with a distinct set of genes. (A) Consensus sequence and secondary structure of subtype 2 *ykkC* motif RNAs. Annotations are as described for **Figure 3-1A**. Asterisks designate the nucleotide positions that are most distinct from the predominant subtype 1 *ykkC* motif RNAs. (B) Genes associated with subtype 2 *ykkC* motif RNAs.

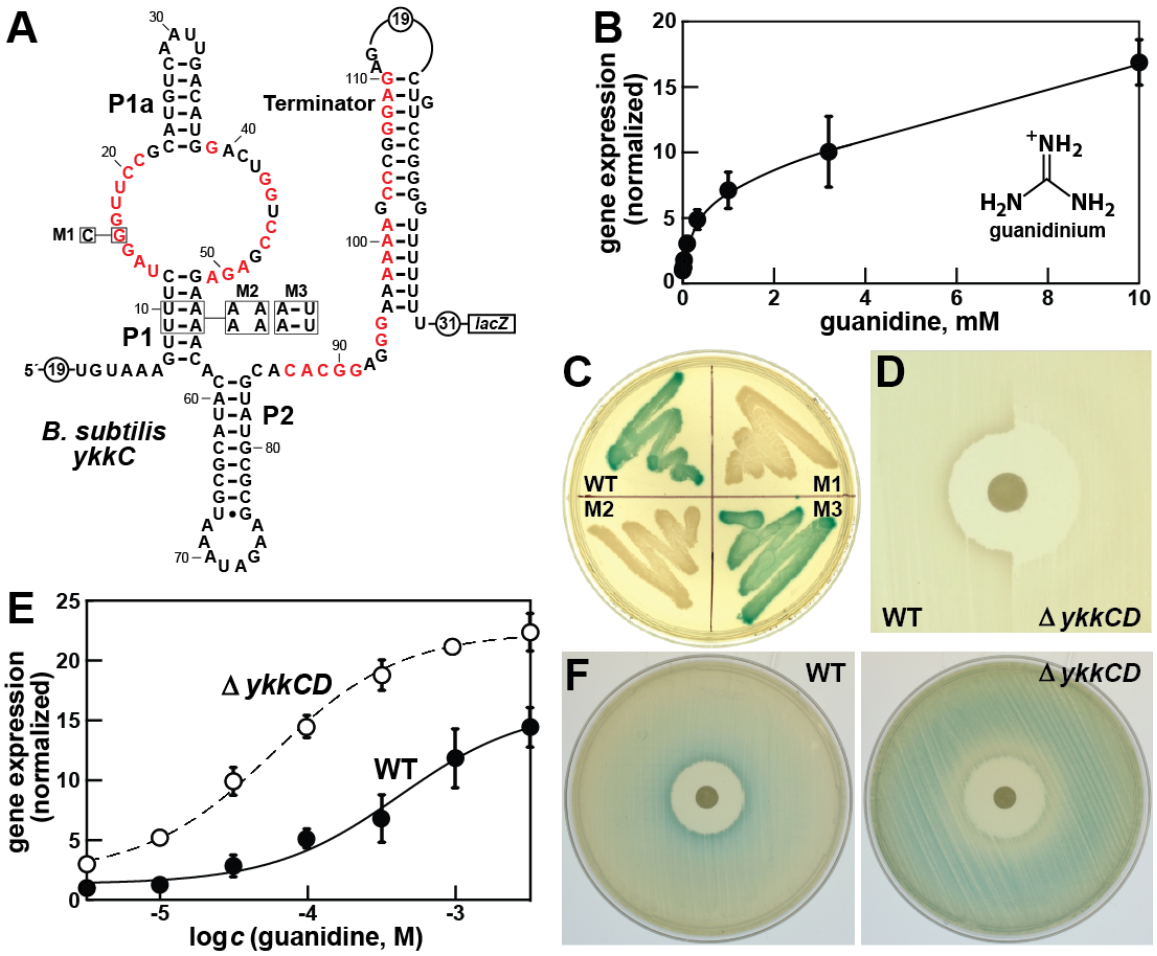

Figure 3-3. A *ykkC* motif RNA responds to guanidine and controls the expression of a transporter that alleviates guanidine toxicity. (A) Sequence and predicted secondary structure of the *ykkC*-reporter construct based on the *ykkC* motif representative associated with the *B. subtilis ykkCD* operon. Red nucleotides are >97% conserved. The secondary structure depicted is proposed to form in the absence of the riboswitch ligand. As with many *ykkC* motif RNAs, part of the predicted aptamer region overlaps a downstream terminator stem such that the presence of the riboswitch ligand will disrupt terminator stem formation, thereby resulting in increased reporter expression. Mutations of conserved nucleotides and secondary structures are designated M1 through M3. (B) Plot of β-galactosidase reporter gene expression in *B. subtilis* versus the concentration of

guanidine, wherein expression in the absence of added guanidine is set to 1. Data points are the average of three measurements and are representative of experiments performed on multiple days. Error bars indicate the standard deviation of the measurements, and when not present are smaller than the data point. Structure for guanidinium is depicted, as this is the predominant form of guanidine under the assay conditions. (C) Reporter gene expression of *B. subtilis* containing wild-type (WT) and mutant reporter constructs as described in A when grown on Lysogeny Broth (LB) agar containing 10 mM guanidine and 100 μg mL^{-1} X-gal. (D) Agar diffusion assay of wild-type and *ykkCD* knockout (Δ*ykkCD*) *B. subtilis* cells grown on the same LB agar plate. Paper disk (center) was spotted with 10 μL of a 6 M guanidine hydrochloride solution (pH 4.7 at 23ºC). (E) Plot of the β-galactosidase expression levels for WT (set to 1 when measured in the lowest guanidine concentration) and Δ*ykkCD B. subtilis* cells at various concentrations of guanidine. Data are the average of three measurements and are representative of experiments performed on multiple days. Error bars represent the standard deviation of the measurement. Other annotations are as described in B. (F) Agar diffusion assays of *B. subtilis* WT and Δ*ykkCD* cells. Disks were spotted with 10 μL of a 6 M guanidine solution. Both plates contained LB medium supplemented with 100 μg mL^{-1} X-gal. Notably for Δ*ykkCD* cells, we observe an inner halo of growing bacteria that do not exhibit reporter gene expression in the presence of high guanidine levels. A similar phenomenon for other bacteria bearing riboswitch reporters in the presence of toxic compounds (Kim et al., 2015).

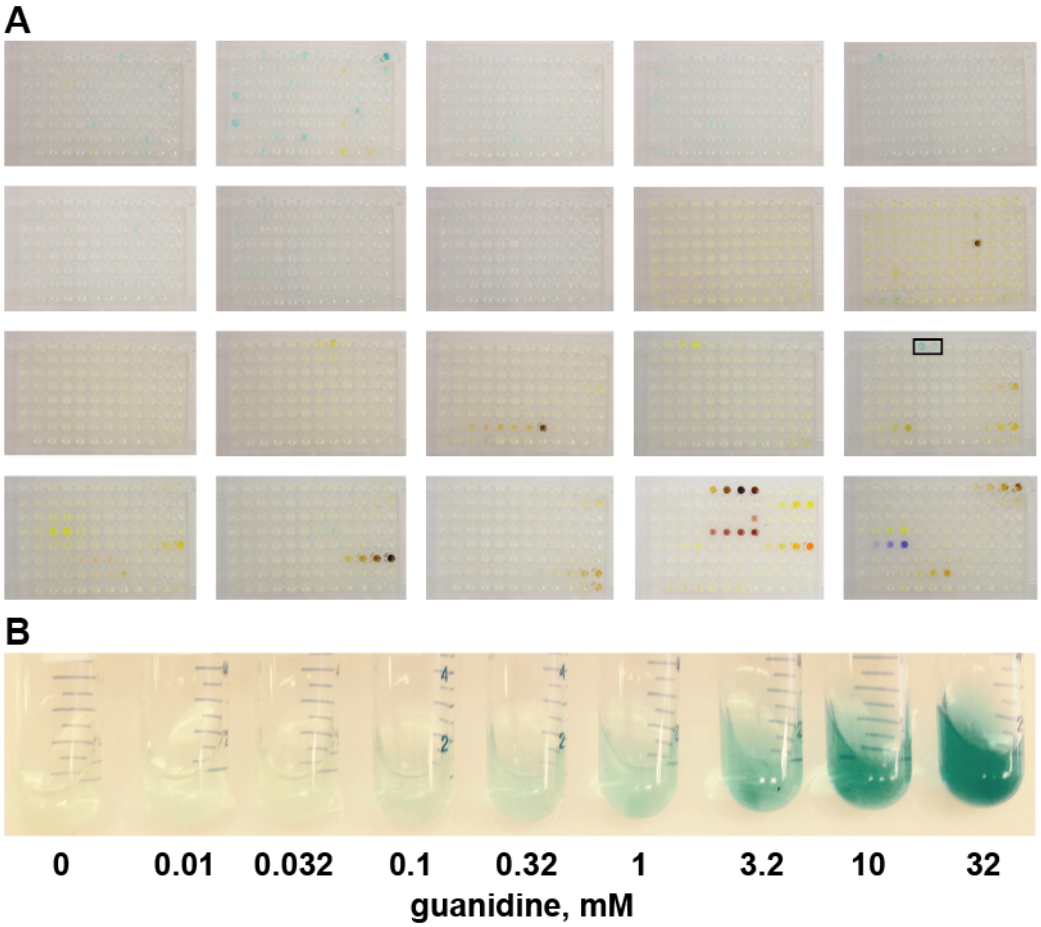

Figure 3-4. Screening growth conditions for activation of a subtype 1 *ykkC* riboswitch reporter fusion yields guanidine as the sole valid hit. (A) Activation of reporter expression in *B. subtilis* by guanidine hydrochloride in Biolog screening assays. Bacteria were grown either in 0.2x LB or proprietary Biolog media supplemented with various compounds and 1 mM X-gal. Two wells containing guanidine are identified by the box. Other conditions that resulted in activation of the *ykkC* riboswitch reporter (light blue color) also activate an azaaromatic riboswitch reporter fusion construct (Li et al., 2016) in the same genetic background, indicating that these conditions likely induce expression of the galactosidase enzyme whose gene is naturally present in the *B. subtilis* genome. Additional colored wells are the result of the inclusion of colored test

compounds in the Biolog screening library. (B) Activation of *ykkC* riboswitch reporter expression in *B. subtilis* by guanidine. Bacteria were grown in 0.2x LB supplemented with the indicated concentration of guanidine and 50 μg mL^{-1} X-gal overnight without shaking.

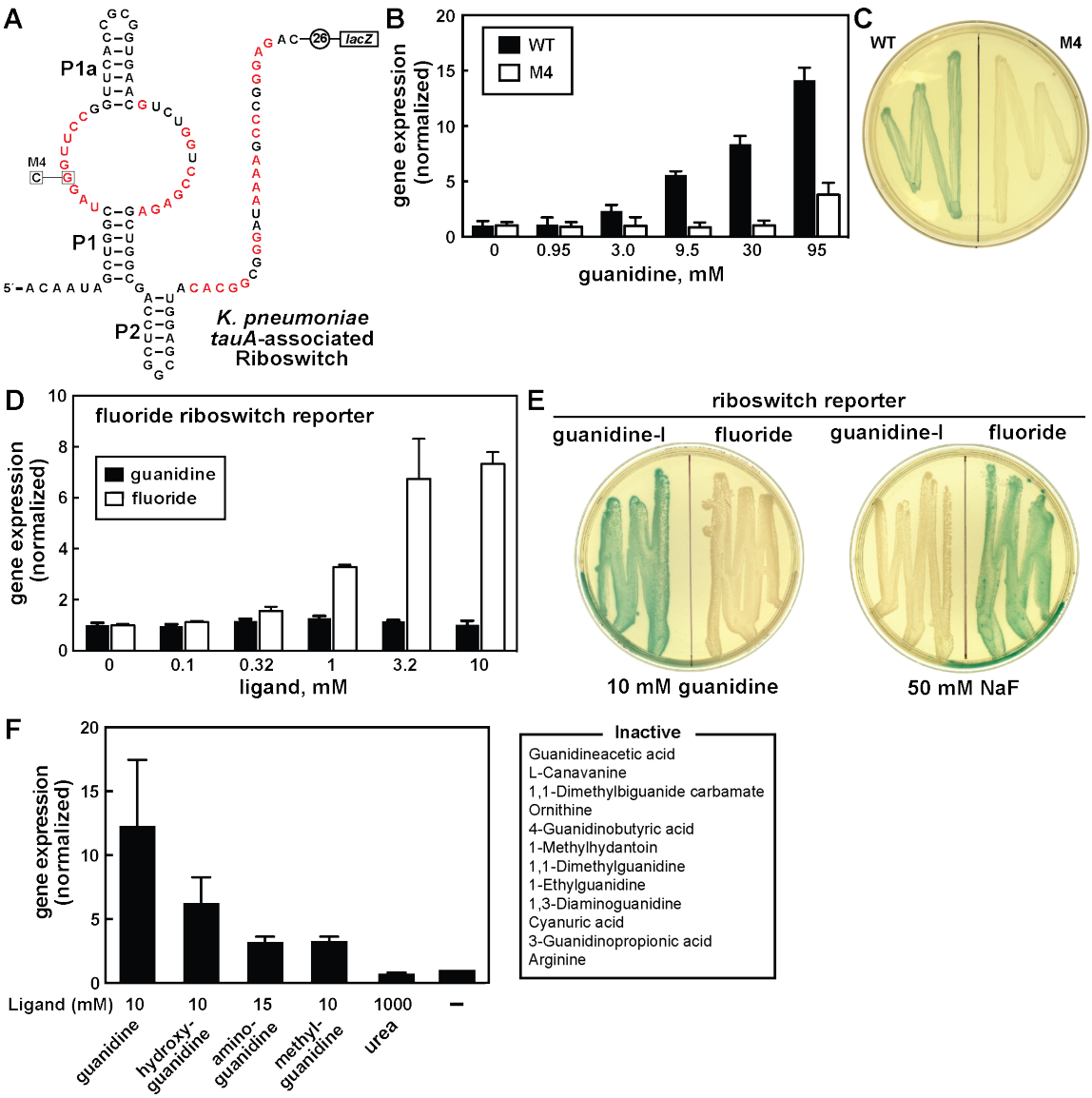

Figure 3-5. Guanidine selectively activates riboswitch reporter constructs. (A) Sequence and secondary structure of a riboswitch reporter construct derived from the subtype 1 *ykkC* motif RNA associated with the *K. pneumoniae tauA* gene (ATPase transporter). Red letters identify the most highly conserved nucleotides within the motif. The expression platform structure and mechanism are not known for this example. M4 designates a mutation at a conserved nucleotide within the riboswitch aptamer. (B) Plot of the β-galactosidase activity for *E. coli* cells containing either the WT or M4 *K.*

pneumoniae riboswitch reporter construct versus the concentration of guanidine. Data are the average of three measurements, and error bars represent the standard deviation of those measurements. Gene expression was normalized to WT *E. coli* grown in the absence of guanidine. (C) LB agar plates containing 100 μg mL^{-1} X-gal and 100 mM guanidine streaked with *E. coli* bearing either the WT or M4 mutant *K. pneumoniae* riboswitch reporter construct as indicated. (D) Plot of β-galactosidase expression of *B. subtilis* cells bearing a fluoride riboswitch reporter (Baker et al., 2012) versus the addition of guanidine or sodium fluoride. Data are the average of three measurements, and error bars represent the standard deviation of those measurements. Gene expression was normalized to the level of β-galactosidase activity measured from host *B. subtilis* cells grown in the absence of guanidine or fluoride. (E) LB agar plates containing 100 μg mL^{-1} X-gal and either 10 mM guanidine or 50 mM sodium fluoride were streaked with *B. subtilis* containing either a guanidine-responsive (**Figure 3-3A**) or fluoride (Baker et al., 2012) riboswitch reporter as indicated. (F) Plot of the relative β-galactosidase expression of *B. subtilis* cells containing the WT guanidine-I riboswitch reporter construct (**Figure 3-3A**) upon addition of various analogs of guanidine at the indicated concentrations using assay conditions as noted for **Figure 3-3**. All other compounds were examined at a concentration either equal to their solubility limit or one-half their MIC. (–): no added compound. Gene expression was normalized to the host *B. subtilis* bacteria grown in the absence of guanidine. Values shown are the average of three replicates. Error bars indicate the standard deviation of the measurement.

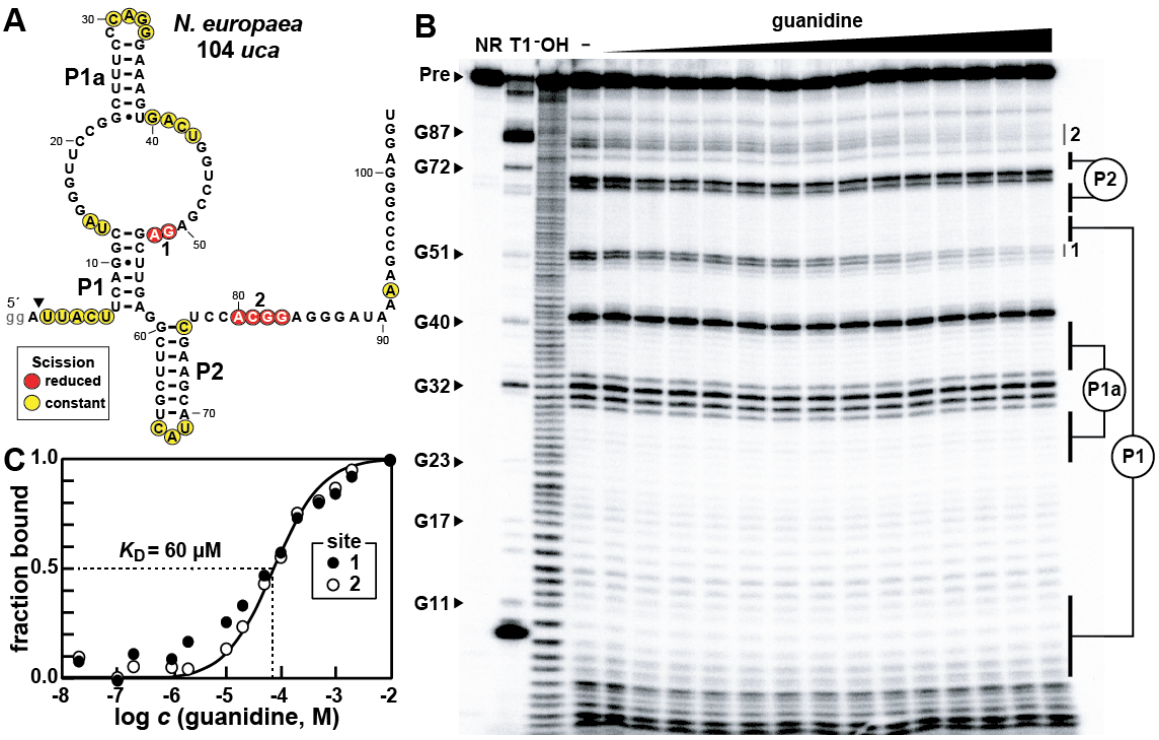

Figure 3-6. Ligand binding by a guanidine-I riboswitch aptamer. (A) Sequence and secondary structure of the 104 *uca* RNA construct derived from the *uca* gene of *N. europaea*. Lowercase letters are nucleotides added to promote efficient *in vitro* transcription with T7 RNA polymerase. Data from the in-line probing assay depicted in B were used to map regions of constant and reduced RNA scission. Arrowhead denotes the start of mapping of available data on the gel. (B) Polyacrylamide gel electrophoresis (PAGE) analysis of the products of in-line probing of 5′ ^{32}P-labeled 104 *uca* RNA in the presence of added guanidine hydrochloride. NR, T1, and $^{-}$OH, respectively, designate RNA undergoing no reaction, or RNAs that have been partially digested with ribonuclease T1 (cleaves after each G), or partially digested under alkaline conditions (cleaves after every nucleotide). Bands corresponding to cleavage after certain G residues in the sequence are annotated. Guanidine concentrations were progressively increased from no ligand added (–) to 10 mM (far right lane). Regions 1 and 2 undergo guanidine-

dependent structure stabilization. (C) Plot of the fraction of RNA bound by guanidine as inferred from the normalized fraction of spontaneous RNA scission at regions 1 and 2 (as depicted in A and B) relative to the logarithm of the molar concentration of guanidine. The solid line is an idealized binding curve expected for a 1-to-1 interaction between a receptor and its ligand. Data shown are representative of multiple experiments.

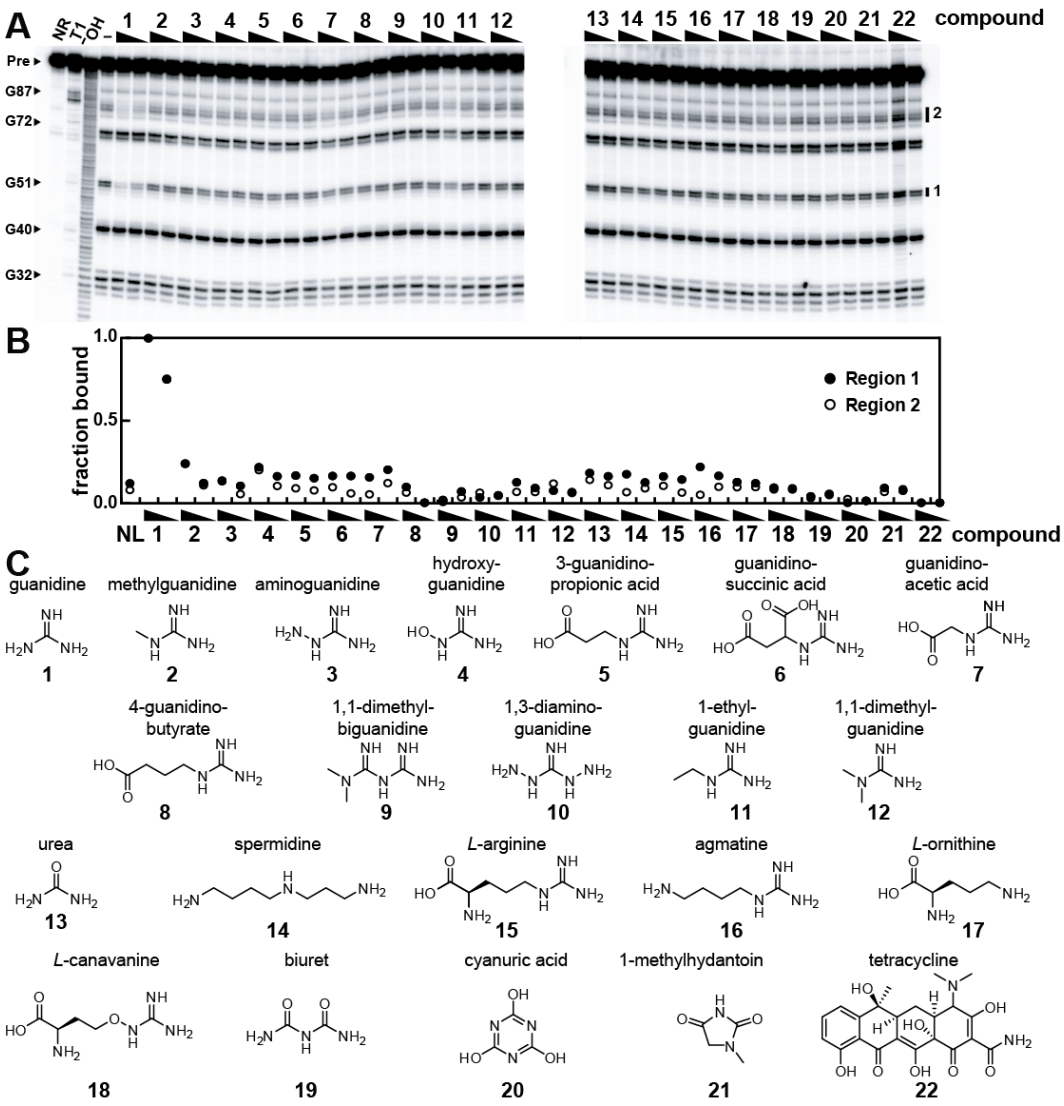

Figure 3-7. Selective recognition of guanidine by a representative guanidine-I riboswitch aptamer. (A) PAGE analysis of in-line probing reactions of the 104 *uca* aptamer from *N. europaea* in the presence of various analogs of guanidine (1 or 0.1 mM). Annotations are as described for **Figure 3-6B**. (B) Plot of the fraction of RNA bound at 1 or 0.1 mM for each compound tested. Data were calculated as described in **Figure 3-6C** using the data presented in A. (C) Chemical structures of compounds tested in A. Although tetracycline (22) is not an analog of guanidine, this compound was examined due to a recent claim that it serves as the ligand for *ykkC* RNAs (Frecker, 2013).

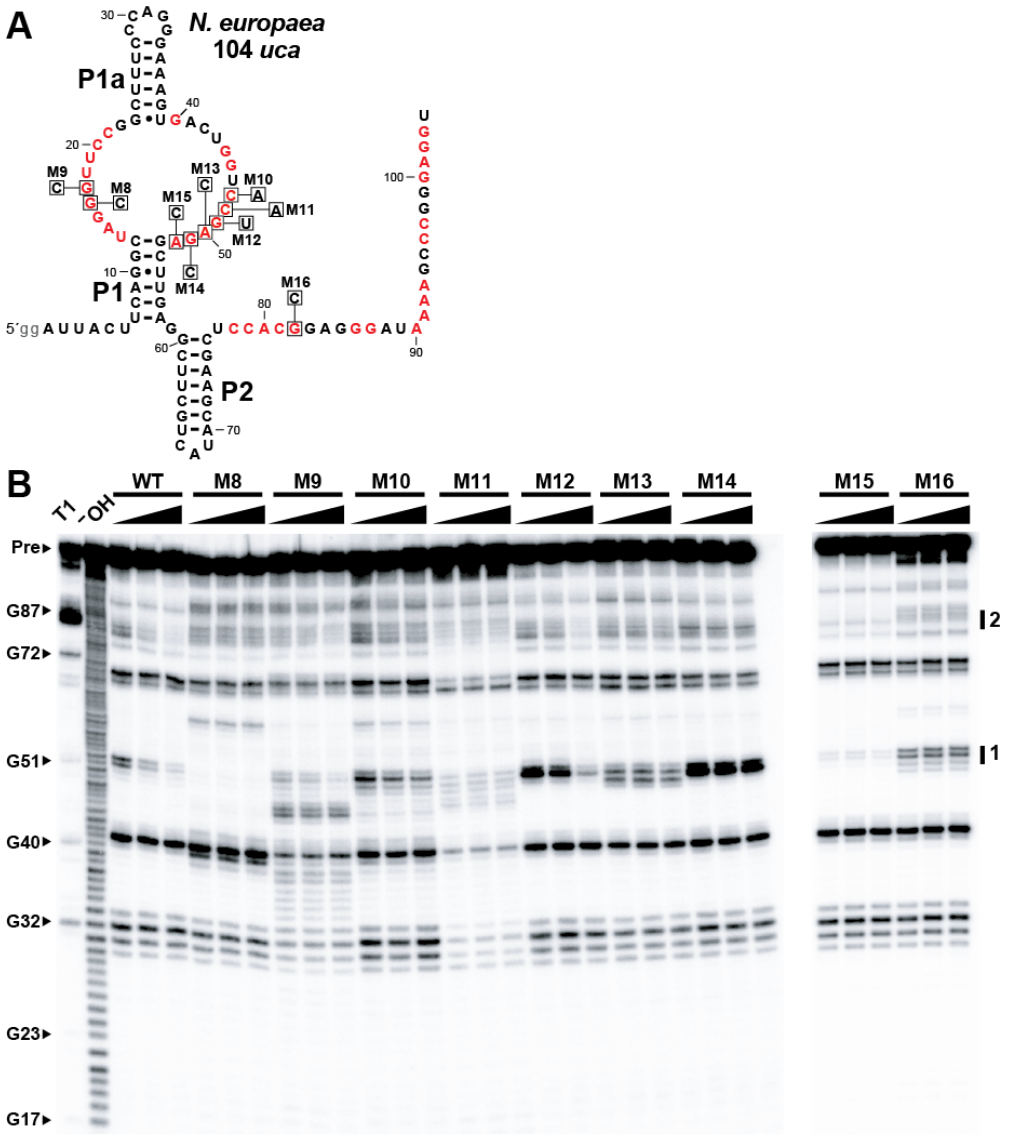

Figure 3-8. Mutation of numerous individual conserved nucleotides disrupts guanidine binding. (A) Location of mutations M8 through M16 within the 104 *uca* RNA construct (**Figure 3-6A**) derived from a guanidine-I riboswitch from *N. europaea*. Annotations are as described for **Figure 3-6A**. (B) PAGE analysis of in-line probing reactions of the 5′ ^{32}P-labeled 104 *uca* RNA from *N. europaea* containing various mutations to conserved nucleotides or secondary structures in the presence of 0, 0.1 mM, or 1 mM guanidine. Additional annotations are as described for **Figure 3-6B**.

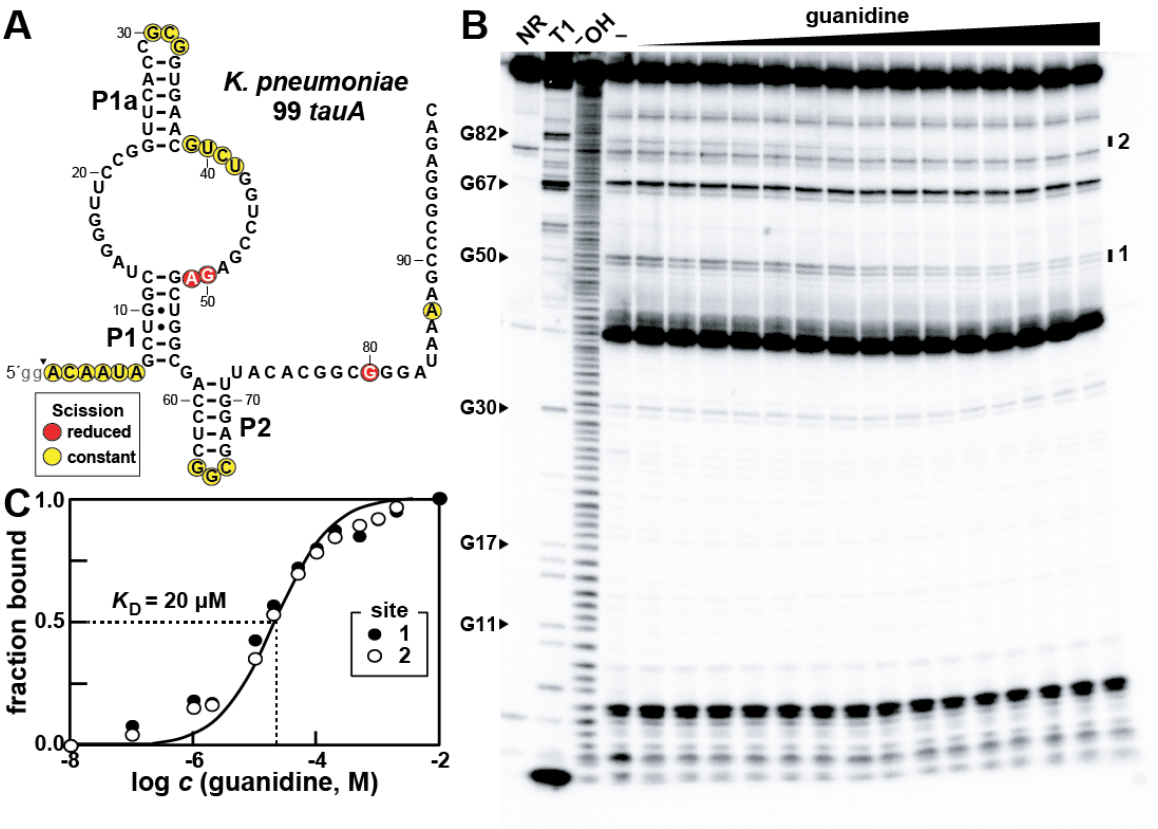

Figure 3-9. Ligand binding by a guanidine-I riboswitch aptamer from K. pneumoniae. (A) Sequence and secondary structure of the 99 *tauA* RNA from *K. pneumoniae*. This is the aptamer domain of the riboswitch used in the reporter system described in **Figure 3-5**. Annotations are as described for **Figure 3-6A**. Sites of ligand-induced structural modulation were determined using the data in B. (B) PAGE analysis of in-line probing reactions with 99 *tauA* in the presence of various guanidine concentrations. Annotations are as described for **Figure 3-6B**. (C) Plot of the fraction of RNA bound by guanidine derived from the normalized fraction of spontaneous RNA scission relative to the logarithm of the molar concentration of guanidine in B. The solid line is an idealized binding curve expected for a 1-to-1 binding interaction. Data are representative of multiple experiments.

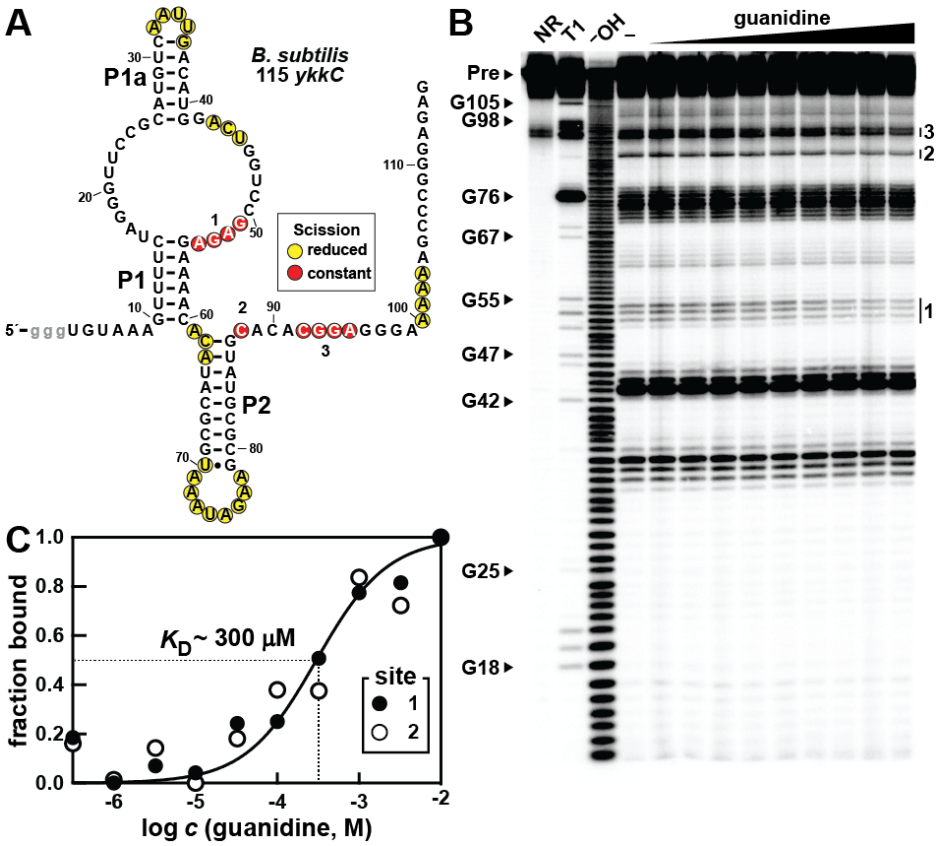

Figure 3-10. Ligand binding by a guanidine-I riboswitch aptamer from *B. subtilis*. (A) Sequence and secondary structure of the 115 *ykkC* RNA from *B. subtilis*. This is the aptamer domain of the riboswitch used in the reporter system described in **Figure 3-3A**. Annotations are as described for **Figure 3-6A**. Sites of ligand-induced structural modulation were determined using the data in B. (B) PAGE analysis of in-line probing reactions with 115 *ykkC* in the presence of various guanidine concentrations. Annotations are as described for **Figure 3-6B**. (C) Plot of the fraction of RNA bound by guanidine derived from the normalized fraction of spontaneous RNA scission relative to the logarithm of the molar concentration of guanidine in B. The solid line is an idealized binding curve expected for a 1-to-1 binding interaction. Data are representative of multiple experiments.

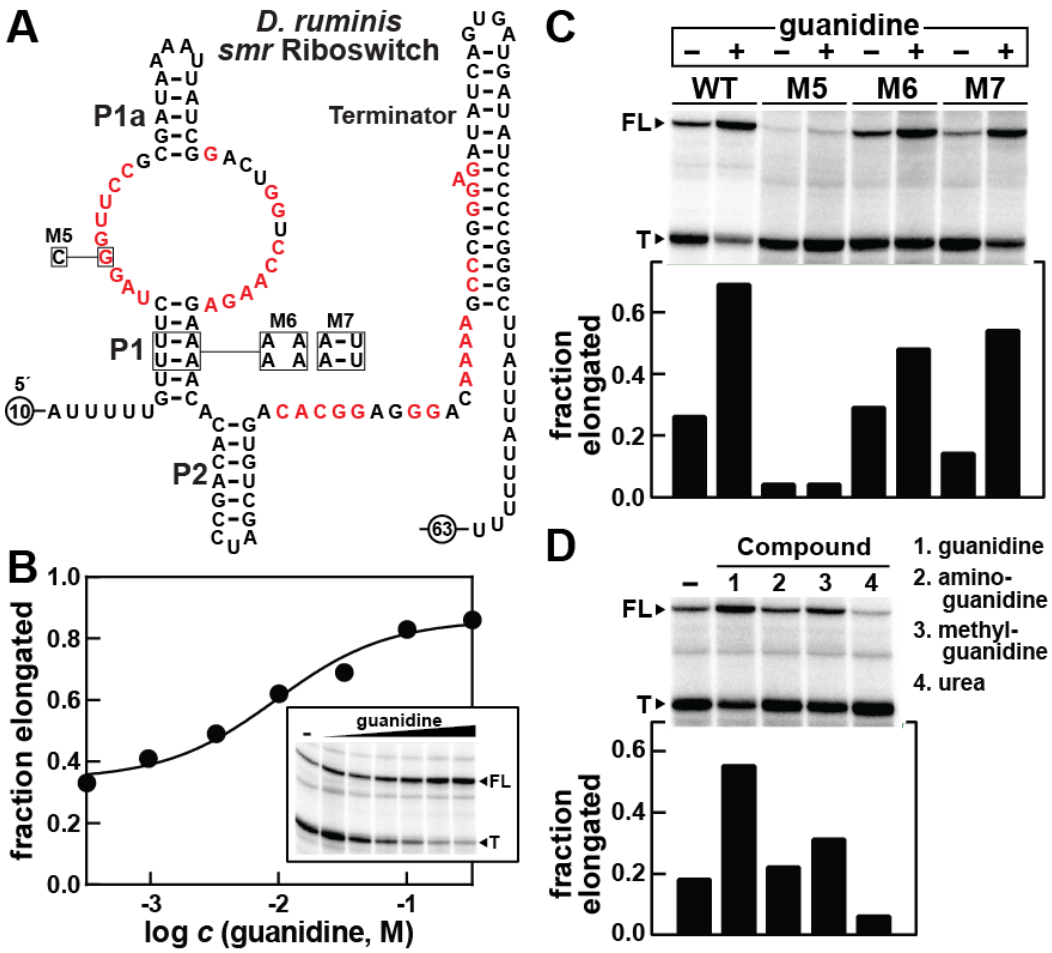

Figure 3-11. Guanidine regulates riboswitch-dependent transcript termination. (A) Sequence and secondary structure model of the guanidine-I riboswitch construct derived from the *smr* gene of *D. ruminis* and used for *in vitro* transcription assays. Annotations are as described in the legend for **Figure 3-3A**. Terminated transcripts end within the U-rich tract immediately following the terminator stem, whereas full-length transcripts include the 63 nucleotides following the U-rich tract. (B) Plot of the fraction of full-length (FL) *smr* riboswitch transcripts contributing to the total number transcripts [FL plus terminated (T)] generated by single-round transcription reactions versus the logarithm of the molar concentration of guanidine. The curve is an idealized representation of the results expected for a 1-to-1 interaction between guanidine and the

riboswitch given the observed minimum and maximum values for fraction elongated. Inset: PAGE analysis used to generate the data plotted. Data shown here and in parts C and D are representative of repeated experiments performed on different days. (C) (Top) PAGE analysis of single-round transcription termination assays of WT and various mutant *smr* riboswitches. (Bottom) Plot of the fraction of FL RNA generated as derived from the PAGE data. (D) (Top) PAGE analysis of single-round transcription termination assays of WT *smr* riboswitches in the absence of ligand (–), in the presence of guanidine or its analogs (10 mM), or in the presence of urea (0.5 M). (Bottom) Plot of the fraction of FL RNA generated as derived from the PAGE data.

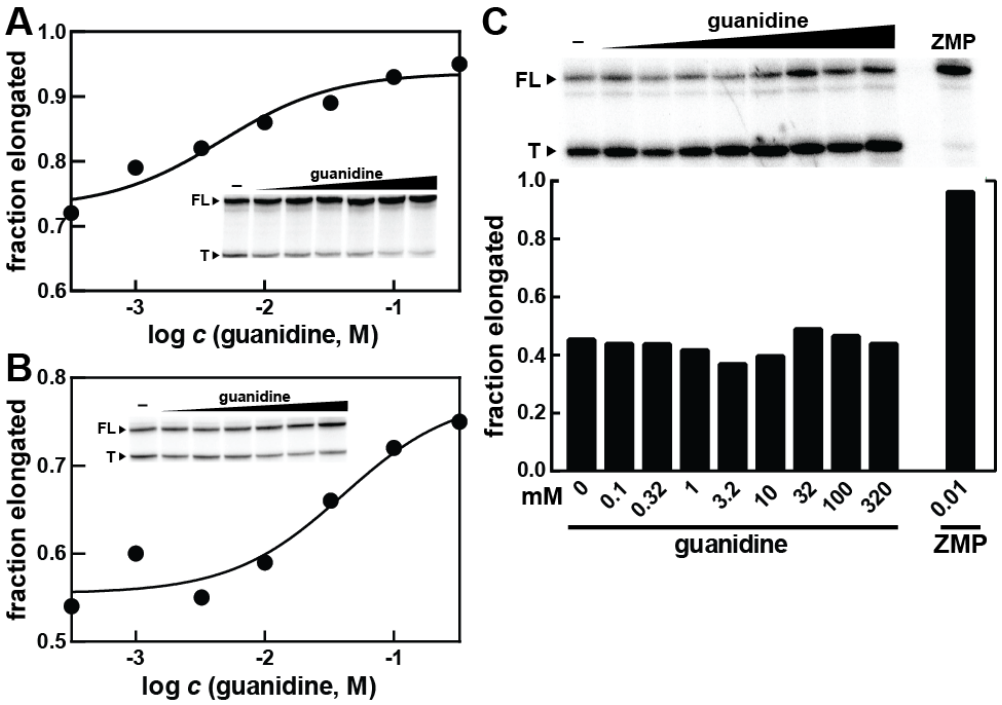

Figure 3-12. Guanidine regulates transcription termination of additional guanidine-I riboswitch representatives, but not a representative from another riboswitch class. (A) Plot of the fraction of FL RNA transcripts in single-round transcription assays of the guanidine-I riboswitch associated with the *emrE* gene from *C. difficile* versus the molar concentration of guanidine. Inset: PAGE analysis used to generate the data plotted. Annotations are as described for **Figure 3-11B**. (B) Plot of the fraction of FL RNA transcripts in single-round transcription assays of the guanidine-I riboswitch associated with the *ykkC* gene from *B. subtilis* versus the molar concentration of guanidine. Inset: PAGE analysis used to generate the data plotted. Annotations are as described for **Figure 3-11B**. (C) (Top) PAGE analysis of single-round transcription assays of the ZTP riboswitch from *Clostridium beijerinckii* (Kim et al., 2015) in the absence (–) or presence of guanidine or ZMP. (Bottom) Plot of the fraction of FL RNA generated as derived from the PAGE data.

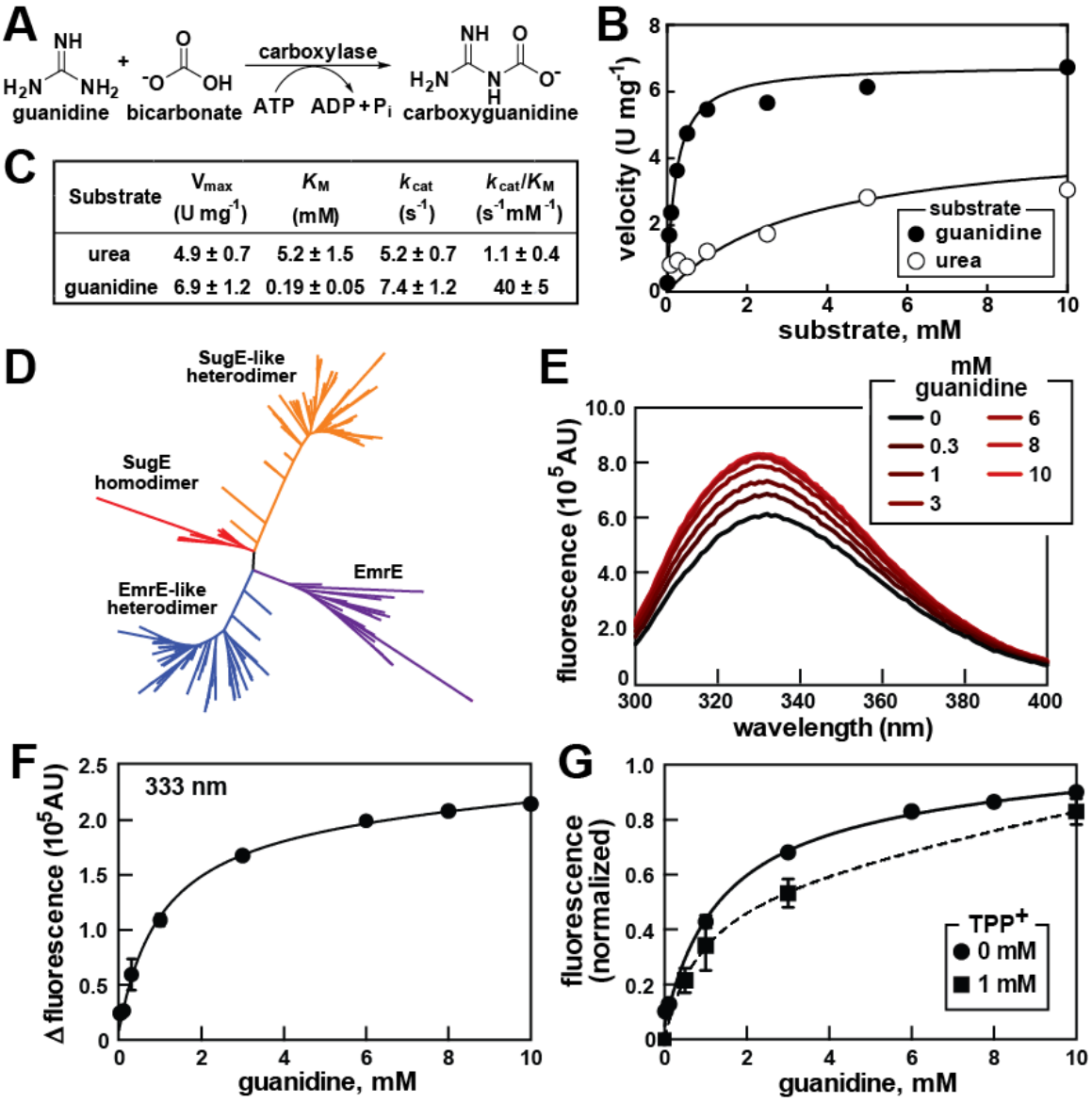

Figure 3-13. Guanidine-related functions for protein products of genes associated with guanidine-I riboswitches. (A) Proposed reaction scheme for the carboxylation of guanidine by an enzyme previously proposed to function as a urea carboxylase (Kanamori et al., 2004). (B) Plot of the velocities of a previously defined urea carboxylase versus the concentration of guanidine or urea. Data points are the average of three technical replicates. (C) List of the enzymatic properties of a carboxylase enzyme with guanidine and urea substrates. Values shown are the averages and standard

deviations of three independent experiments. (D) Phylogenetic tree showing the genetic variability within SMR transporters controlled by guanidine-I riboswitches. Selected experimentally validated EmrE proteins that are known not to be controlled by the riboswitch are shown in purple. (E) Recognition of guanidine by predicted SMR transporters controlled by guanidine riboswitches evaluated by tryptophan emission spectra of *Clo* SugE at pH 7.5 in the presence of increasing guanidine. (F) Plot of the increase in fluorescence of tryptophan residues within *Clo* SugE at 333 nm versus the concentration of guanidine. Background fluorescence in the absence of guanidine is subtracted from each data point. Solid line represents the fit to a 1-to-1 binding model with a K_D of 1 mM. Error bars indicate the standard error of the experiment. When not visible, error bars are smaller than the data point. (G) Plot of the intrinsic fluorescence of tryptophan residues within *Clo* SugE at 333 nm versus the concentration of guanidine in the presence and absence of 1 mM TPP^+. Other annotations are as described for F. Data are the average of three measurements. Error bars indicate the standard error of the experiment. When not visible, error bars are smaller than the data point.

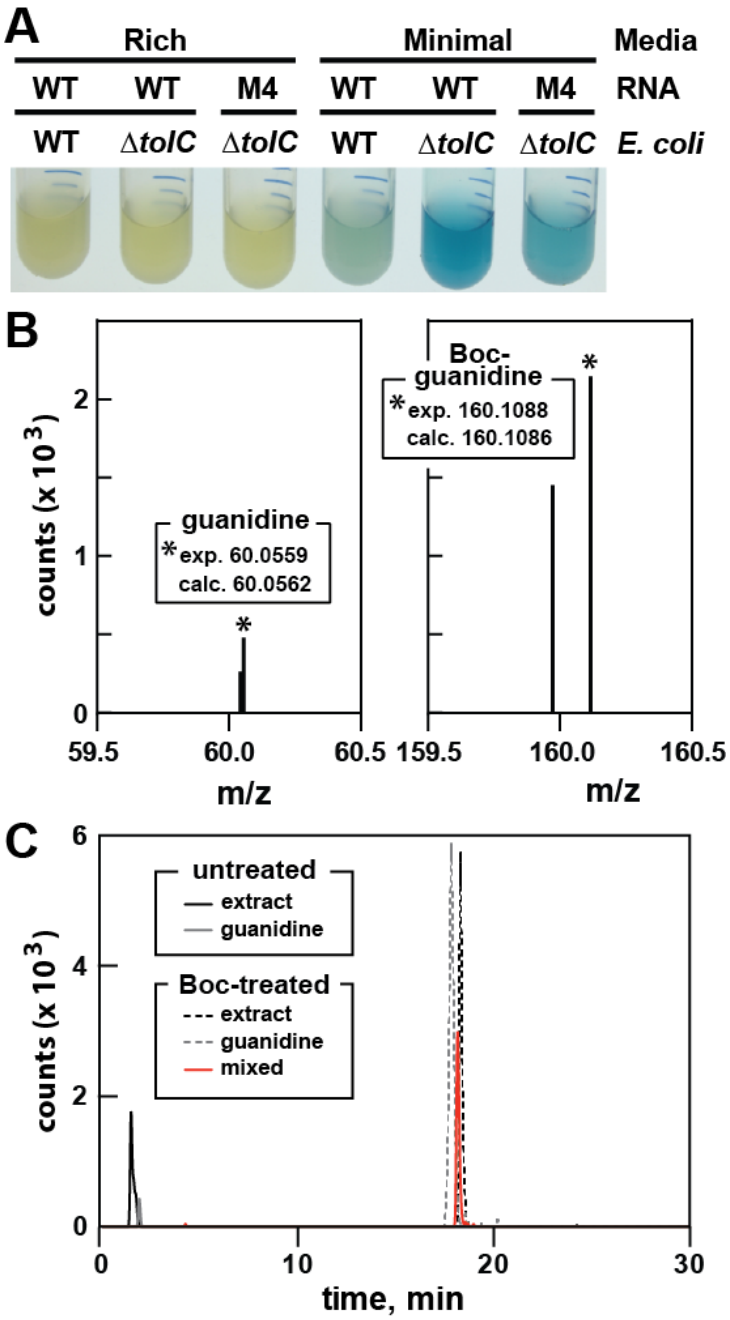

Figure 3-14. Natural production of free guanidine by bacteria. (A) *E. coli* WT and *tolC* knockout (Δ*tolC*) cells carrying the *K. pneumoniae* WT or M4 riboswitch reporter constructs (**Figure 3-5**) grown in rich (LB) or minimal (GMM) liquid media. (B) Plots of the mass-to-charge ratios (m/z) of compounds with similar retention times to authentic guanidine and authentic Boc-guanidine. Asterisks identify peaks from the extract LC-MS

spectrum within 10 ppm of the calculated mass of each authentic compound as designated. (C) LC-MS analysis for guanidine from the Δ*tolC* strain of *E. coli* grown in minimal medium as depicted in A. Extract and authentic guanidine samples were analyzed directly or were treated with Boc-anhydride to alter the mobility and mass of free guanidine as noted.

Table 3-1. Sequences of synthetic DNAs used in Chapter 3.

Name	Sequence	Purpose
ykkC001	TAATACGACTCACTATAGGGTGTAAAGTTTTCTAGGGTTCCGCATGTCAATTGACATGGACTGGTCCGAGAGAAAAC	Forward primer for the construction of a template for transcription of the *B. subtilis* 100 *ykkC* RNA
ykkC002	CTCTCCCGGGCTTTTTTCCCTCCGTGTGCATACGCGCTTCTATTTACGCGTATGTGTTTTCTCTCGGACCAGTCC	Reverse primer for the construction of a template for transcription of the *B. subtilis* 100 *ykkC* RNA
ykkC003	TAATACGACTCACTATAGGACAATAGCTGGCTAGGGTTCCGGTTCACCGCGGTGAACGTCTGGTCCGAGAGCTGGCGAC	Forward primer for the construction of a template for transcription of the *K. pneumoniae* 99 *tauA* RNA
ykkC004	GTCTCCCGGGCTTTTATCCCGCCGTGTAACCTCGCCGAGGTCGCCAGCTCTCGGACCAG	Reverse primer for the construction of a template for transcription of the *K. pneumoniae* 99 *tauA* RNA
ykkC005	TAATACGACTCACTATAGGATTACTTCAGGCTAGGGTTCCGGCTTTCCCAGGGAAAGTGACTGGTCCG	Forward primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA
ykkC006	ACCTCCCGGGCTTTTATCCCTCCGTGGAGCTTCGTATGACGAAGCCTCAAGCTCTCGGACCAGTCACTTTCCCTG	Reverse primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA
ykkC007	TAATACGACTCACTATAGGATTACTTCAGGCTAGCGTTCCGGCTTTCCCAGGGAAAGTGACTGGTCCG	Forward primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M8
ykkC008	TAATACGACTCACTATAGGATTACTTCAGGCTAGGCTTCCGGCTTTCCCAGGGAAAGTGACTGGTCCG	Forward primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M9
ykkC009	TAATACGACTCACTATAGGATTACTTCAGGCTAGGGTTCCGGCTTTCCCAGGGAAAGTGACTGGTACG	Forward primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M10
ykkC010	ACCTCCCGGGCTTTTATCCCTCC	Reverse primer for the construction of a

	GTGGAGCTTCGTATGACGAAGCC TCAAGCTCTCGTACCAGTCACTT TCCCTG	template for transcription of the *N. europaea* 104 *uca* RNA containing M10
ykkC011	TAATACGACTCACTATAGGATTA CTTCAGGCTAGGGTTCCGGCTTT CCCAGGGAAAGTGACTGGTCAG	Forward primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M11
ykkC012	ACCTCCCGGGCTTTTATCCCTCC GTGGAGCTTCGTATGACGAAGCC TCAAGCTCTCTGACCAGTCACTT TCCCTG	Reverse primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M11
ykkC013	TAATACGACTCACTATAGGATTA CTTCAGGCTAGGGTTCCGGCTTT CCCAGGGAAAGTGACTGGTCCT	Forward primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M12
ykkC014	ACCTCCCGGGCTTTTATCCCTCC GTGGAGCTTCGTATGACGAAGCC TCAAGCTCTAGGACCAGTCACTT TCCCTG	Reverse primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M12
ykkC015	ACCTCCCGGGCTTTTATCCCTCC GTGGAGCTTCGTATGACGAAGCC TCAAGCTCGCGGACCAGTCACTT TCCCTG	Reverse primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M13
ykkC016	ACCTCCCGGGCTTTTATCCCTCC GTGGAGCTTCGTATGACGAAGCC TCAAGCTGTCGGACCAGTCACTT TCCCTG	Reverse primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M14
ykkC017	ACCTCCCGGGCTTTTATCCCTCC GTGGAGCTTCGTATGACGAAGCC TCAAGCGCTCGGACCAGTCACTT TCCCTG	Reverse primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M15
ykkC018	ACCTCCCGGGCTTTTATCCCTCG GTGGAGCTTCGTATGACGAAGCC TCAAGCTCTCGGACCAGTCACTT TCCCTG	Reverse primer for the construction of a template for transcription of the *N. europaea* 104 *uca* RNA containing M16
ykkC019	TAATACGACTCACTATAGGCAAT GATATGCTATAAAATAATCTCCG GACGAGACGCGCCGTACCCGGT	Forward template for transcription of the *L. pneumophila* 71 *ykkC*–III RNA

	TAAGACAAGACGACGCCTTTA	
ykkC020	TAAAGGCGTCGTCTTGTCTTAAC CGGGTACGGCGCGTCTCGTCCGG AGATTATTTTATAGCATATCATT GCCTATAGTGAGTCGTATTA	Reverse template for transcription of the *L. pneumophila* 71 *ykkC*–III RNA
ykkC021	ATAAGTGTAAAATTATTCTTATT GACAAACTAAATGGTTTGGTATT ATTCTTATATTAGATATGATGAT ATCATGTAAAGTTTTCTAGGGTT CCGCATGTCAATTGACATGGACT GGTCCGAGAGAAAACACATACG CGTAAATAGAAGCGCGTATG	Forward primer for the construction of single-round transcription termination template from *B. subtilis*
ykkC022	TTTGAGTCATTCCTTTCTATCTAT TAAGATTCTCCGAAAACAAAAA AACCCCGGACAGTCGTCTCACAT GAGATTGACTCTCCCGGGCTTTT TTCCCTCCGTGTGCATACGCGCT TCTATTTACGCGTATGTGTTTTCT CTCGG	Reverse primer for the construction of single-round transcription termination template from *B. subtilis*
ykkC023	ATAAGTGTAAAATTATTCTTATT GACAAACTAAATGGTTTGGTATT ATTCTTATATTAGATAGAATAAT TAATAGAGTTTTATAGGGTTCCG CAATTATTAATGATTGGTTTGGT CCGAGATAAAACCCACAACATA GTTTTTGTTGTGAACACGGAAGG	Forward primer for the construction of single-round transcription termination template from *C. difficile*
ykkC024	TTTGAGTCATTCCTTTCTATCTAT TAAGATTCTCCGATCCCTCTTTTC AAATATGTCTCTTATTTTTCCAAT AAAATAACCTGAGAAATATTAT AAAATATCTCCCAGGTTTTTGTC CTTCCGTGTTCACAACAAAAACT ATGTTG	Reverse primer for the construction of single-round transcription termination template from *C. difficile*
ykkC025	ATAAGTGTAAAATTATTCTTATT GACAAACTAAATGGTTTGGTATT ATTCTTATATTAGATAAAATATA	Forward primer for the construction of single-round transcription termination template from *D. ruminis*

	TATTTTTGTTTTCTAGGGTTCCGC GATAAAATTATCGGACTGGTCCA AGAGAAAACACACAGCCTAGCT GTGACACGGAGGGACAAA	
ykkC026	TTTGAGTCATTCCTTTCTATCTAT TAAGATTCTCCGGAAAGTCATGT CTACTTCCTCCTTATTAAAAAAT AAATAAGCCCGGGGATATCATC ACTGATATCTCCCGGGCTTTTGT CCCTCCGTGTCACAGCTAGGCTG TGTGTTTT	Reverse primer for the construction of single-round transcription termination template from *D. ruminis*
ykkC027	TTTGAGTCATTCCTTTCTATCTAT TAAGATTCTCCGGAAAGTCATGT CTACTTCCTCCTTATTAAAAAAT AAATAAGCGGTTTTATATCATCA CTGATATCTCCCGGGCTTTTGTC CCTCCGTGTCACAGCTAGGCTGT GTGTTTT	Reverse primer for the construction of single-round transcription termination template from *D. ruminis* containing a mutation in the terminator stem
ykkC028	ATAAGTGTAAAATTATTCTTATT GACAAACTAAATGGTTTGGTATT ATTCTTATATTAGATAAAATATA TATTTTTGTTTTCTAGCGTTCCGC GATAAAATTATCGGACTGGTCCA AGAGAAAACACACAGCCTAGCT GTGACACGGAGGGACAAA	Forward primer for the construction of single-round transcription termination template from *D. ruminis* with a mutation at a conserved nucleotide
ykkC029	ATAAGTGTAAAATTATTCTTATT GACAAACTAAATGGTTTGGTATT ATTCTTATATTAGATAAAATATA TATTTTTGTAATCTAGGGTTCCG CGATAAAATTATCGGACTGGTCC AAGAGAAAACACACAGCCTAGC TGTGACACGGAGGGACAAA	Forward primer for the construction of single-round transcription termination template from *D. ruminis* with a mutation in the P1 stem
ykkC030	ATAAGTGTAAAATTATTCTTATT GACAAACTAAATGGTTTGGTATT ATTCTTATATTAGATAAAATATA TATTTTTGTAATCTAGGGTTCCG	Forward primer for the construction of single-round transcription termination template from *D. ruminis* with a compensatory mutation in the P1 stem

	CGATAAAATTATCGGACTGGTCCAAGAGATTACACACAGCCTAGCTGTGACACGGAGGGACAAA	
ykkC031	ATAAGTGTAAAATTATTCTTATTGACA	Forward primer for the amplification of all termination constructs
ykkC032	TTTGAGTCATTCCTTTCTATC	Reverse primer for the amplification of all termination constructs
ykkC033	TACGACGAATTCCAAAAATAATG	Forward primer for the construction of *B. subtilis ykkC* reporter construct containing the *B. subtilis lysC* promoter containing sites for cloning into pDG1661
ykkC034	TCCTTTGGATCCTATTAAGATTCTCCG	Reverse primer for the construction of *B. subtilis ykkC* reporter construct containing the *B. subtilis lysC* promoter, containing sites for cloning into pDG1661
ykkC035	TACGACAAATTGCAAAAATAATGTTGTCCTTTTAAATAAGATCTGATAAAATGTGAACTAATATGATGATATCATGTAAAGTTTTCTA	Forward primer for the construction of *B. subtilis ykkC* termination construct containing the *B. subtilis lysC* promoter, used as a template for construction of the *B. subtilis ykkC* reporter construct
ykkC036	TTTGGAGTCATTCCTTTCTATCTATTAAG	Reverse primer for the construction of *B. subtilis ykkC* termination construct containing the *B. subtilis lysC* promoter, used as a template for construction of the *B. subtilis ykkC* reporter construct
ykkC037	GATGATATCATGTAAAGTAATCTAGG	Forward primer for construction of the *B. subtilis ykkC* reporter with the *lysC* promoter containing M2 via QuikChange mutagenesis
ykkC038	CCTAGATTACTTTACATGATATCATC	Reverse primer for construction of the *B. subtilis ykkC* reporter with the *lysC* promoter containing M2 via QuikChange mutagenesis
ykkC039	CATGTAAAGTTTTCTAGCGTTC	Forward primer for construction of the *B. subtilis ykkC* reporter with the *lysC* promoter containing M1 via QuikChange

		mutagenesis
ykkC040	GAACGCTAGAAAACTTTACATG	Reverse primer for construction of the *B. subtilis ykkC* reporter with the *lysC* promoter containing M1 via QuikChange mutagenesis
ykkC041	GGTCCGAGAGATTACACATACG	Forward primer for construction of the *B. subtilis ykkC* reporter with the *lysC* promoter containing M3 via QuikChange mutagenesis
ykkC042	CGTATGTGTAATCTCTCGGACC	Reverse primer for construction of the *B. subtilis ykkC* reporter with the *lysC* promoter containing M3 via QuikChange mutagenesis
ykkC043	CTGCCAGGAATTGGGGATCGGA ATTCGCATGACCCCATCGTTGAC AACCGCCCCGCTCACCCTTTATT TATAAATGTACCAATTCAGATCA CACAATAGCTGGCTAGGGTTCCG GTTCACCGCGGTGAACGTCTGGT CCGAGAGCTGGCGACCTCGGCG AGGTTACACGGCGGGATAAAAG CCCGGGAGACAGCAGCACCGTT GGGTGTCGCGCTGCTCCGTTTCT CGTGCTAACCAGAGGTCCACCAT GACCGTGATTACGGATTCAGGG GATCCCGTCGTTTTACAACGTCG	IDT G-block containing the wild-type *K. pneumoniae tauA* riboswitch and first 8 codons of *lacZ* gene controlled by the *E. coli lysC* promoter, for cloning into pRS414
ykkC044	CTGGCTAGCGTTCCGGTTC	Forward primer for construction of the *K. pneumoniae tauA* reporter with the *lysC* promoter containing M4 via QuikChange mutagenesis
ykkC045	GAACCGGAACGCTAGCCAG	Reverse primer for construction of the *K. pneumoniae tauA* reporter with the *lysC* promoter containing M4 via QuikChange mutagenesis
ykkC046	TTGGTGATGAGCTCGTTAGGAGG GGAACTGAACATGAAG	Forward primer for amplification of the 5' flank of the *ykkCD* KO cassette, for

		cloning into pDG1515 and subsequent construction of the knockout via homologous recombination
ykkC047	ATCACCAACTGCAGCTTTCAATTGTAAGATTTTCTTTCATATACCGC	Reverse primer for amplification of the 5' flank of the *ykkCD* KO cassette, for cloning into pDG1515 and subsequent construction of the knockout via homologous recombination
ykkC048	TTGGTGATGTCGACGCTCAGCAGTTGGTTTAAAAATTCTGTC	Forward primer for amplification of the 3' flank of the *ykkCD* KO cassette, for cloning into pDG1515 and subsequent construction of the knockout via homologous recombination
ykkC049	ATCACCAAGGTACCGGGCCTCTTGCAGTCTAGTTAAAGAC	Reverse primer for amplification of the 3' flank of the *ykkCD* KO cassette, for cloning into pDG1515 and subsequent construction of the knockout via homologous recombination

Chapter Four

A bacterial riboswitch class for the thiamin precursor HMP-PP employs a terminator-embedded aptamer

Largely adapted from a manuscript being prepared for publication:

Atilho RM, Mirihana Arachchilage G, Greenlee EB, Knecht KM, Breaker RR. A bacterial riboswitch class for the thiamin precursor HMP-PP employs a terminator-embedded aptamer. *eLife* (Submitted*)*

Summary

My colleague Shira Stav implemented a bioinformatics pipeline that can uncover novel, but rare, riboswitch candidates as well as other noncoding RNA structures in bacteria. A prominent candidate revealed by her initial search efforts was called the '*thiS* motif' because of its frequent association with a gene coding for the ThiS protein, which delivers sulfur to form the thiazole moiety of the thiamin precursor HET-P. In Chapter 4, my colleagues and I describe biochemical and genetic data demonstrating that *thiS* motif RNAs function as sensors of the thiamin precursor HMP-PP, which is fused with HET-P ultimately to form the final active coenzyme thiamin pyrophosphate (TPP). HMP-PP riboswitches exhibit a distinctive architecture wherein an unusually small ligand-sensing aptamer is almost entirely embedded within an otherwise classic intrinsic transcription terminator stem. This arrangement yields remarkably compact genetic switches that bacteria use to tune the levels of thiamin precursors during the biosynthesis of this universally distributed coenzyme.

Introduction

Approximately 40 distinct riboswitch classes that regulate gene expression in various bacterial species have been experimentally validated to date (McCown et al., 2017; Serganov and Nudler, 2013; Sherwood and Henkin, 2016; Breaker, 2011a). Based on the abundances and distributions of these known riboswitch classes, it has been proposed that many thousands of additional riboswitch classes remain to be discovered in the eubacterial domain of life (Ames and Breaker, 2010; Breaker, 2011b; McCown et al., 2017). The collection of known riboswitch classes largely sense compounds or ions that are of fundamental importance to organisms from all three domains of life, and these ligands also exhibit a bias in favor of compounds (enzyme cofactors, RNA nucleotides and their precursors or derivatives) that are predicted to be of ancient origin (Breaker, 2011b; McCown et al., 2017; Nelson and Breaker, 2017). If these trends hold, it seems likely that numerous additional riboswitch classes that regulate fundamental biological processes remain to be discovered. Unfortunately, the vast majority of these undiscovered riboswitch classes are predicted to be exceedingly rare, and this characteristic is likely to cause difficulties for researchers who seek to identify them.

To address this challenge, Shira Stav developed a computational pipeline that first identifies the regions of bacterial genomes that are most likely to serve as transcription templates for structured noncoding RNAs (ncRNAs), and then uses comparative sequence and structural analyses to identify novel candidate RNA motifs (Stav et al., 2019). Specifically, this approach examines only the putative noncoding regions of a given sequenced bacterial genome, and evaluates each intergenic region (IGR) based on two parameters: (*i*) percent guanosine and cytidine (GC) nucleotide content and (*ii*)

length in nucleotides. For many bacteria, structured ncRNAs are GC-rich compared to other regions of the genome (Klein et al., 2002; Schattner, 2002), and the IGRs that serve as synthesis templates for these ncRNAs tend to be much longer than typical bacterial IGRs that contain only an RNA polymerase promoter and/or a protein-specific regulatory domain. This current bioinformatics pipeline is based on earlier implementations of this search strategy that were used to discover several novel structured ncRNA motifs (Meyer et al., 2009), including the SAM-V riboswitch class (Poiata et al., 2009).

Shira Stav's updated computational pipeline was employed to comprehensively examine the genomes of five bacterial species, which revealed the existence of as many as 70 novel genetic elements, including 30 candidate ncRNA motifs (Stav et al., 2019). Of the eight candidate riboswitch classes uncovered in this search, the most promising was called the '*thiS* motif' (**Figure 4-1A**) because representatives commonly reside immediately upstream of *thiS* genes, which code for a protein that delivers sulfur to the pathway for the production of the thiamin biosynthetic intermediate 5-(2-hydroxyethyl)-4-methylthiazole phosphate (HET-P) (Begley et al., 2012). Furthermore, *thiS* motif RNAs associate with additional genes for enzymes that catalyze the synthesis of HET-P or its fusion to 4-amino-5-hydroxymethyl-2-methylpyrimidine diphosphate (HMP-PP), to ultimately produce the bioactive coenzyme thiamin pyrophosphate (TPP) (Jurgenson et al., 2009).

An additional clue regarding the function of *thiS* motif RNAs was derived from the fact that this novel RNA structure occasionally resides in tandem with TPP riboswitches (Stav et al., 2019). Tandem arrangements of other riboswitch classes have been shown to function as two-input Boolean logic gates (Sudarsan et al., 2006; Stoddard and Batey,

2006; Lee et al., 2010; Sherlock et al., 2018), suggesting that the genes associated with tandem arrangements of TPP riboswitches and *thiS* motif RNAs likely respond to concentration changes of two distinct ligands. Together, these observations strongly indicate that *thiS* motif RNAs function as riboswitches that respond to a biochemical intermediate of the TPP biosynthetic pathway.

Previously (Stav et al., 2019), I created a riboswitch-reporter fusion construct by joining a *thiS* motif RNA representative to a β-galactosidase gene. Using this construct, I observed robust gene expression in host *Bacillus subtilis* cells with a deleted *thiS* gene (Δ *thiS*). This same reporter-fusion construct yields no reporter gene expression in host *B. subtilis* cells that naturally carry a *thiS* gene. These findings were consistent with our hypothesis that *thiS* motif RNAs function as riboswitches, but the precise ligand sensed by the unusual architecture of this RNA remained unknown.

In Chapter 4, I described the bioinformatic, genetic and biochemical analyses that provided conclusive evidence that the ligand for *thiS* motif RNAs is the TPP biosynthetic intermediate HMP-PP. Furthermore, these findings demonstrate that this unusually small riboswitch employs a distinct architecture to regulate RNA transcription termination. This regulatory RNA provides cells with an efficient mechanism to balance the production of two key biosynthetic intermediates, HMP-PP and HET-P, which are then fused to make the essential coenzyme TPP.

Results

Architecture and genetic distribution of the *thiS* riboswitch candidate

Most experimentally validated riboswitch classes are composed of distinct, but partially overlapping aptamer and expression platform domains (Barrick and Breaker, 2007; Breaker, 2011b). In contrast, the *thiS* motif exhibits an unusual arrangement wherein the predominant secondary structure, derived by thermodynamic modeling, is an extended hairpin structure that maximizes conventional Watson/Crick base pairing (**Figure 4-1A**, left). This structure exhibits all the features characteristic of bacterial intrinsic terminator stems, including an uninterrupted and strong base-paired stem followed by a run of six or more uridine (U) residues (Wilson and von Hippel, 1995; Yarnell and Roberts, 1999). As a result, we concluded that transcription termination was certain to be a major function of *thiS* motif RNAs.

We frequently encounter simple terminator stems when using bioinformatics search algorithms to identify novel ncRNA candidates, and we have now begun to assign predicted terminator function to such motifs and then quickly move on to examine other more promising ncRNA candidates. When evaluating the *thiS* candidate, however, we noted four features that suggested this terminator stem was peculiar. First, the loop of the terminator hairpin is abnormally well conserved compared to the loop sequences of more typical terminator stems, which are usually irrelevant to the mechanism of transcription termination. This unique terminator element is found in species from two phyla and from several classes within Firmicutes. Second, Etienne Greenlee demonstrated another consensus structural model was also consistent with the comparative sequence analysis data (**Figure 4-1A**, right). This architecture, including a possible pseudoknot and two

major base-paired stems, disrupts the contiguous terminator stem near the run of U residues, suggesting that mutually exclusive and competing structures and functions might exist for *thiS* motif RNAs. Third, these apparently specialized terminator stems associate exclusively with genes related to the biosynthesis and utilization of HET-P, which is a precursor of the coenzyme TPP (**Figure 4-1B**). These sequence, structure, and genomic distribution characteristics suggested to us that each *thiS* motif RNA might function as a compact ligand sensor and regulator of TPP coenzyme biosynthesis.

A fourth feature of *thiS* motif RNAs is that approximately 30% of the known representatives reside immediately downstream of riboswitches that sense and respond to TPP (Stav et al., 2019). Each associated TPP riboswitch appears to use a terminator stem as an expression platform. The tandem arrangement of two terminator stems for a single riboswitch aptamer would be unprecedented, and so this observation also supported our hypothesis that *thiS* motif RNAs represent an unusual form of riboswitch. Due to these tandem arrangements, we speculated that the natural ligand for this riboswitch candidate would not be TPP. Rather, it seemed more likely that the ligand would be a major precursor of TPP (either HET-P or HMP-PP), and that each tandem TPP-*thiS* motif system would function as a two-input Boolean logic gate (Sudarsan et al., 2006; Stoddard and Batey, 2006; Lee et al., 2010; Sherlock et al., 2018) to regulate HET-P production in response to the cellular concentrations of both TPP and one of its biosynthetic precursors.

Furthermore, if the two secondary structure states called the ‘terminator configuration’ and the ‘aptamer configuration’ (**Figure 4-1A**) represent the riboswitch ‘OFF’ and ‘ON’ configurations, respectively, then ligand binding by the aptamer in *thiS* motif RNAs is expected to activate gene expression. The vast majority of riboswitches

that sense metabolites and control biosynthetic pathways are OFF switches, and thus the accumulation of the metabolite ligand results in reduced expression of the protein products that otherwise would make (or import) more of the desired metabolite. Although *thiS* motif representatives associate with genes for the production of HET-P (**Figure 4-1B**), it does not make sense that HET-P would turn on its own production when it is already abundant. Finally, I recognized that the *thiE* gene (coding for thiamin-phosphate synthase) is occasionally associated with *thiS* motif representatives, suggesting that the RNA motif determines if cellular conditions are suitable to couple the two key precursors of thiamin monophosphate (TMP) to eventually yield TPP. Taken together, all these bioinformatic observations are consistent with the hypothesis that *thiS* motif RNAs function as riboswitches for the TPP precursor HMP-PP.

A *thiS* riboswitch-reporter fusion construct is activated by the addition of HMP

As an initial test of our hypothesis that HMP-PP is the ligand for *thiS* motif RNAs, I employed a chromosomally-integrated reporter construct (Stav et al., 2019) carrying a β-galactosidase gene fused downstream of nucleotides encompassing the *thiS* motif from *Clostridium* species Maddingley (Rosewarne et al., 2013) (**Figure 4-2A**). Transformed *B. subtilis* cells carrying a transcriptional fusion of the wild-type (WT) riboswitch to a *lacZ* reporter gene, which was integrated into the *amyE* locus, exhibited no β-galactosidase activity in response to HMP added to rich (LB) liquid culture media (**Figure 4-2B**). Perhaps the natural suppression of TPP biosynthesis when cells have sufficient amounts of this coenzyme precludes the formation of excess HMP-PP. Specifically, if translation of the *thiD* gene coding for HMP/HMP-P kinase is largely suppressed under normal

cellular conditions, the externally supplied HMP cannot be phosphorylated to generate HMP-PP.

In contrast, robust β-galactosidase activity is evident in cells co-transformed with the reporter construct and a vector overexpressing the ThiD protein, but only when growth media is supplemented with HMP (**Figure 4-2B**). Reporter gene expression levels are dependent on the concentration of HMP added to the culture medium (**Figure 4-2C**), suggesting that cells take up HMP and use the *thiD* gene product to add two phosphates to generate excess HMP-PP, which activates *lacZ* reporter expression by binding to the riboswitch.

Furthermore, the characteristics of a series of mutant riboswitch-reporter constructs examined in *B. subtilis* cells likewise indicates that *thiS* motif RNAs function as riboswitch aptamers for HMP-PP. Constructs carrying mutations M1 (G26C) or M2 (C35G) (**Figure 4-2A**), which alter strictly-conserved nucleotides that are predicted to be part of the pseudoknot structure of the aptamer (**Figure 4-1A**), are not activated by HMP addition (**Figure 4-2C**). Importantly, construct M3 (U46A, U47A, U48A) exhibits a detectable level of reporter gene expression in LB media alone, and this expression is further enhanced by the addition of HMP. This result suggests that these nucleotides are involved in forming a strong terminator stem, but that the identities of the nucleotides at these positions are not critical for ligand binding by the putative HMP-PP aptamer.

I also used the WT riboswitch-reporter construct (**Figure 4-2A**) to assess gene regulation in response to differences in growth media, and to several genetic disruptions of the TPP biosynthetic pathway (**Figure 4-3**). *B. subtilis* cells grown in LB medium are expected to suppress the genes needed to produce both the pyrimidine moiety HMP-PP

and the thiazole moiety HET-P. Thus, as expected, expression of the reporter gene fused to the putative HMP-PP riboswitch is off, regardless of the genetic background tested. Likewise, WT cells grown in minimal (GMM) medium also do not express the reporter gene, presumably because HMP-PP does not accumulate due to its rapid and efficient conversion to the final product, TPP. By contrast, Δ*thiS* and Δ*thiE* cells (carrying genetic disruptions of the *thiS* and *thiE* genes, respectively) exhibit high levels of reporter gene expression (**Figure 4-3**). These two genetic knock-out strains are predicted to accumulate HMP-PP because they either lack the protein (ThiS) that initiates the production of HET-P, or they lack the protein (ThiE) that fuses HMP-PP to HET-P even when HET-P is available.

Additional support for the hypothesis that HMP-PP is the ligand, rather than HMP-P, is provided by the level of expression of the reporter construct when present in host cells carrying a deletion of the gene (Δ*thiD*) required to phosphorylate HMP-P to make HMP-PP. Reporter gene expression is not observed in Δ*thiD* cells (**Figure 4-3C**), despite the fact that these cells should accumulate HMP-P. Overall, my genetic results strongly indicate that HMP-PP is the ligand for a riboswitch class represented by *thiS* motif RNAs that turns on gene expression when this ligand is abundant.

HMP-PP suppresses intrinsic transcription termination *in vitro*

The two possible architectures of the *thiS* motif RNA from *C.* sp. Maddingley (**Figure 4-4A**), also observed in other members of this candidate riboswitch class (**Figure 4-1A**), suggest a mechanism for gene regulation involving the mutually exclusive formation of an intrinsic terminator stem and its competing ligand-bound aptamer state. Such

transcription control mechanisms for riboswitches have been experimentally validated in the past (e.g. McDaniel et al., 2003; Mironov et al., 2002; Sudarsan et al., 2003; Wickiser et al., 2005a) by using single-round *in vitro* transcription assays (Landick et al., 1996) to reveal ligand-dependent modulation of RNA transcript lengths. Therefore, I sought to examine the proposed riboswitch mechanism by conducting in vitro transcription reactions with the expected natural ligand, HMP-PP.

Unfortunately, my use of such transcription assays was made more difficult, because HMP-PP is not commercially available and is known to be relatively unstable (Hanes et al., 2007). Therefore, I freshly generated HMP-PP from HMP and ATP enzymatically (**Figure 4-4B**) by using recombinantly produced ThiD protein from Kirsten Knecht. HMP-PP production as described previously (Hanes et al., 2007) was confirmed by mass spectrum analysis (**Figure 4-4C**). The enzymatic reactions were deproteinized by filtration, and the resulting mixtures containing HMP-PP were added, without further purification, to various assays as described for each experiment.

The DNA template for the WT *thiS* motif RNA construct (**Figure 4-4A**) yields only ~14% full-length transcript when transcription reactions are conducted in the presence of HMP alone. In contrast, nearly 40% of the transcripts are full-length (FL) when transcription reactions are conducted in the presence of enzyme-prepared HMP-PP (**Figure 4-4D**). Furthermore, mutations M1 and M2, which disrupt possible pseudoknot formation by the aptamer and which eliminate gene expression in vivo (**Figure 4-2C**), fail to produce more FL transcripts when HMP-PP is present. My results suggest that HMP-PP binding induces RNA polymerase to transcribe past the intrinsic terminator element to yield fewer terminated (T) products and a greater proportion of FL RNAs.

From the assay data presented, this switching effect is incomplete in vitro, as evident by the fact that construct M4, which carries three nucleotide changes that disrupt the otherwise perfect base pairing of the terminator stem, yields nearly 100% 'FL' transcripts. Construct M5 yields transcripts that approximate the length of 'T' RNAs, which are generated naturally by termination due to the action of the intrinsic terminator stem. Incomplete switching by terminator-regulating riboswitches is typical for such assays conducted in vitro (e.g. McDaniel et al., 2003; Mironov et al., 2002; Sherlock et al., 2018; Sudarsan et al., 2003; Wickiser et al., 2005a). However, it seems likely that a larger dynamic range for transcription control is exploited by this riboswitch in cells, as evident by the robust differences in gene expression for the nearly identical reporter constructs used in this study (**Figure 4-2, Figure 4-3**).

Biochemical evidence for direct binding of HMP-PP by RNA aptamers

The unique architecture of *thiS* motif RNAs also served as an obstacle for evaluating the ability of these RNAs to directly bind a ligand. My colleagues and I frequently employ in-line probing assays (Soukup and Breaker, 1999; Regulski and Breaker, 2008) to determine if RNAs undergo structural changes in response to ligand binding. For *thiS* motif RNAs, the terminator stem is expected to dominate over the aptamer configuration (**Figure 4-1A**) simply due to the different number of base-pairs present in each structural state. Therefore, under thermodynamic equilibrium conditions typically experienced by RNAs subjected to in-line probing reactions, a full-length construct is not expected to bind HMP-PP because the RNA will always favor the terminator configuration ('OFF' state).

To address this problem, I reasoned that a shorter construct that weakens the terminator stem might permit the aptamer configuration to be adopted. Such constructs are also likely to better represent the structures naturally adopted by the riboswitch during transcription. An RNA polymerase paused at the run of U nucleotides at the end of the intrinsic terminator element will sequester ~12 nucleotides of the transcript within the protein structure (Monforte et al., 1990; Komissarova and Kashlev, 1998; Vassylyev et al., 2002). Therefore, only the first 50 to 54 nucleotides of the nascent transcript of the natural *C*. sp. Maddingley *thiS* motif should be exposed if the RNA polymerase complex is paused within the run of U nucleotides of the intrinsic terminator stem.

A series of four *thiS* motif RNA constructs was prepared by Gayan Mirihana Arachchilage to investigate the hypothesis that shorter constructs might permit aptamer formation by weakening the terminator stem. The longest RNA construct carries the full terminator stem (66 *thiS*), and three progressively truncated variants (54 *thiS*, 53 *thiS* and 52 *thiS*) represent only the RNA regions that would be exposed when RNA polymerase is stalled at various locations within the run of U nucleotides (**Figure 4-5A**). In-line probing assays (**Figure 4-5B**) reveal that 66 *thiS* indeed exclusively forms the terminator stem, as evident by robust spontaneous RNA cleavage in the unstructured regions near the 5′ terminus through nucleotide position 9, and by the unstructured loop region including nucleotides 35 through 39. This loop region remains unstructured with all constructs tested, demonstrating that the upper base-paired portion is common to all structural configurations.

Construct 54 *thiS* appears to retain all base-pairs of the terminator stem, except those that are disrupted by the deletion of the 3′-terminal nucleotides from position 54 and

beyond (**Figure 4-5B**). In stark contrast, construct 52 *thiS* adopts a configuration that matches the consensus model for the aptamer configuration, including the formation of stems P1 and P2 (**Figure 4-1A**). Intriguingly, this shortest construct exhibits modest evidence (**Figure 4-5B**, asterisk) of a weakening of base-pairing within the series of G-U wobble interactions formed between nucleotides 24 to 27 and nucleotides 46 to 49. This structural flexibility might permit the formation of a pseudoknot or some other tertiary interactions between these highly-conserved G nucleotides and pyrimidines in the loop region upon ligand binding. Also noteworthy is the fact that the intermediate length construct, 53 *thiS*, exhibits an in-line probing pattern consistent with the formation of both the terminator configuration and the aptamer configuration. These results suggest that the genetic decision whether to form the terminator stem configuration or the ligand-bound aptamer configuration is made within a very narrow window of transcription progression.

Given the ability of the 52 *thiS* construct to exclusively adopt the aptamer configuration, we used this RNA to seek additional evidence for direct binding of HMP-PP by the riboswitch. However, this construct still failed to exhibit ligand-induced changes in the banding pattern resulting from in-line probing reactions. My colleagues and I speculated that nucleotide positions 25-27 and positions 46-48 form base-pairs that trap this portion of the motif in its terminator configuration. To further favor the formation of the desired aptamer configuration, Gayan Mirihana Arachchilage made the M3 version of the 52 *thiS* construct (**Figure 4-6A**). The same M3 mutations, made in the context of the full-length construct, earlier were observed to retain ligand responsiveness in riboswitch-reporter assays in vivo (**Figure 4-2B**).

The 52 *thiS* M3 construct indeed exhibits structural modulation upon introduction of HMP-PP (**Figure 4-6B**). Although the main base-paired regions P1 and P2 remain unchanged by HMP-PP addition, nucleotides involved in forming the putative pseudoknot (sites 1 and 2), and other nucleotides in the loop of P2 (site 3) appear to become more structured. Gayan Mirihana Arachchilage also introduced additional mutations into construct 52 *thiS* M3 to evaluate the effects of mutations known or expected to disrupt gene control function. Specifically, previous mutations M1 or M2 (**Figure 4-2A**), or M6 (**Figure 4-7A**) were introduced to create constructs, termed M7, M8 and M9, which carry alterations to highly-conserved nucleotides that presumably disrupt pseudoknot formation. These additional mutations eliminate structural modulation by HMP-PP (**Figure 4-7B**), as would be expected if these highly-conserved sequence and structural features are critical for riboswitch aptamer function.

We suspect that the ligand concentration is insufficient to saturate all RNAs in the sample. Regardless, by quantifying the band intensities at the three sites of structural modulation and by assuming that these values could become zero (or background) upon ligand binding, we generated a partial HMP-PP binding curve for 52 *thiS* M3 (**Figure 4-6C**), which is consistent with a 1-to-1 interaction between the ligand and the RNA aptamer. Given our use of enzymatically prepared HMP-PP samples, we cannot use this data to precisely determine the dissociation constant (K_D) for this interaction. However, if we assume that all HMP was enzymatically converted into HMP-PP, this product was fully recovered after removal of protein by filtration, and that there was no degradation over the time frame of the assays, the K_D value cannot be greater than 500 μM for this variant aptamer construct.

A series of shortened constructs carrying the M3 mutations (**Figure 4-8A**) recapitulate the transition between the terminator and aptamer configurations as originally observed for the series of WT truncated RNAs (**Figure 4-5**). Importantly, the 53 *thiS* M3 construct exhibits evidence of both binding to HMP-PP and of structural switching to favor the aptamer configuration ('ON' state) (**Figure 4-8B**). Gayan Mirihana Arachchilage's findings suggest that it is necessary to prevent the formation of G-U wobble base-pairs between nucleotide positions 25-27 and positions 46-48 for ligand binding to be observed by using in-line probing assays. The time scale of in-line probing reactions (24 to 48 hours) is much longer than the time scale for the natural genetic decision to take place (probably a few seconds). Therefore our biochemical assays permit the RNAs to reach thermodynamic equilibrium (Wickiser et al., 2005a; Wickiser et al., 2005b), which should favor the terminator configuration. By mutating the U nucleotides at positions 46-48, we prevent constructs from becoming thermodynamically trapped in the terminator configuration during in-line probing, which permits HMP-PP binding to be observed. This same effect could be achieved with natural *thiS* motif sequences by having the U nucleotides at positions 46-48 remain sequestered in the RNA-exit channel (Vassylyev et al., 2002; Hein et al., 2014) of an RNA polymerase that is paused within the run of U nucleotides of the intrinsic terminator stem. We speculate that momentary ligand binding during this paused state will be sufficient to prevent terminator formation on the vastly shorter time scale that is relevant to the genetic decision process in cells.

Overall, our biochemical data are strongly consistent with the hypothesis that *thiS* motif RNAs function as riboswitches that directly bind to HMP-PP. As a result, we favor renaming *thiS* motif RNAs as HMP-PP riboswitches. The pursuit of more precise

biochemical and biophysical characteristics, and further evidence for the mechanism proposed above, will require both the judicious use of RNA constructs that form the aptamer configuration, and either the preparation of pure samples of HMP-PP with consideration for its relative instability or the use of more stable analogs that can trigger riboswitch function.

Tandem TPP and HMP-PP riboswitches function as Boolean genetic logic gates

As noted above, approximately 30% of the representatives of HMP-PP riboswitches reside in tandem with TPP riboswitches. In these tandem systems, the TPP riboswitch always occurs first, and the TPP aptamer is always associated with its own expression platform, which is routinely a readily recognizable intrinsic terminator stem. The HMP-PP riboswitch follows each complete TPP riboswitch, and by its unique architecture, carries its own intrinsic terminator expression platform.

This arrangement, as observed in a representative from the bacterium *Clostridium lundense* (**Figure 4-9A**), makes apparent the genetic decisions made by the host. Specifically, abundant TPP should repress expression of the associated gene, which usually codes for a protein needed to biosynthesize HET-P (**Figure 4-1B**). This repression, regardless of the concentration of HMP-PP, makes sense because cells do not need to make more TPP when this enzyme cofactor is already abundant. However, when TPP is in short supply, transcripts should read through the first intrinsic terminator stem to begin transcription of the HMP-PP riboswitch. The HMP-PP aptamer only permits transcription read-through of its co-resident intrinsic terminator stem if the relative abundance of its ligand is high. If HMP-PP is abundant, then the resulting full-length

mRNAs will produce proteins that biosynthesize more HET-P. This two-input tandem riboswitch system has a truth table that matches a 'converse nonimplication' Boolean logic function (**Figure 4-9B**).

To confirm that a tandem TPP and HMP-PP riboswitch system functions with this Boolean logic, I created a tandem riboswitch-reporter fusion construct based on the natural *C. lundense* RNA (**Figure 4-9A**), and examined its function in *B. subtilis* cells. Unfortunately, I could not easily create cellular conditions to fully examine all four possible ligand states represented in the truth table (**Figure 4-9B**). Most obviously, TPP is an essential cofactor, and thus we cannot deplete its concentration to zero while maintaining cell viability. Also, the various TPP biosynthetic intermediates are likely to be present in low concentrations if the cell is continuously making more TPP, even at low flux. As a result, a cell might experience conditions where an individual nascent RNA transcript of the tandem riboswitch system could respond to any of the four possible states, but these responses when summed over multiple nascent transcripts will yield a hybrid gene expression response that generally reflects the state of ligands in the cell.

To fully examine the gene expression outputs for all four possible states, despite the complications described above, I only partially manipulated the state of ligands, while also creating artificial representations of the missing states by using riboswitch aptamer mutations. Specifically, by adding HMP to the culture medium, I could change the ligand state from state 2 (+ TPP, – HMP-PP) to state 4 (+ TPP, + HMP-PP). As predicted by the truth table, ligand states 2 and 4 do not yield expression of the reporter gene (**Figure 4-9C**). I also artificially created states 1 (– TPP, – HMP-PP) and 3 (– TPP, + HMP-PP) by employing disruptive mutations in the aptamers for TPP (M10), HMP-PP (M1), or both

(M11). Only the artificial state 3 condition, as created by mutating the TPP aptamer and growing cells in the presence of HMP, yields robust reporter gene expression (**Figure 4-9C**, M10). This gene expression effect is lost when disabling mutations are placed in both aptamers (M11) to simulate state 1. These findings are consistent with the hypothesis that each tandem TPP and HMP-PP riboswitch system functions as a Boolean converse nonimplication logic gate.

Discussion

All of our findings derived from bioinformatic, genetic and biochemical analyses indicate that *thiS* motif RNAs represent an unusual class of riboswitches that sense and respond to the thiamin pyrophosphate biosynthetic intermediate HMP-PP. Its architecture is unusual because the majority of the nucleotides and structures required for ligand recognition appear to be encompassed by an intrinsic transcription terminator, which forms in a mutually-exclusive manner to regulate gene expression. All other riboswitch classes reported to date (McCown et al., 2017) that employ an intrinsic terminator structure as an expression platform carry the most-highly conserved nucleotides and structural features largely upstream of, and only partly overlapping, the terminator structure. This unique overlapping architecture of HMP-PP riboswitches provides bacteria with a highly-compact, yet effective RNA-based gene control system for measuring HMP-PP concentrations and tuning the production of HET-P to more efficiently produce TPP.

The embedded nature of the aptamer and expression platform domains, however, complicated the pursuit of experiments typically used to biochemically validate a newly discovered riboswitch class. Most importantly, our experiments required the use of non-native constructs to successfully reveal HMP-PP binding in vitro. Therefore, it is notable that modifications to certain key constructs maintain function in vitro or in vivo. For example, the trimming of nucleotides at the 3′ region of the terminator stem to create RNA construct 52 *thiS* yields a molecule that is expected to represent the exposed portion of a polymerase-paused nascent RNA transcript (**Figure 4-5A**). This truncated construct adopts the predicted P1 stem of the aptamer configuration (**Figure 4-5B**). Also, the addition of nucleotide changes to create the M3 variant of 52 *thiS* yields a construct that

binds HMP-PP, whereas this same M3 mutation in a related construct retains ligand-dependent gene-control function in cells (**Figure 4-2C**).

Our results also reveal a mechanism for gene control involving the stalling of RNA polymerase at a precise location to yield an exposed nascent RNA transcript that is transiently capable of forming an HMP-PP binding pocket. The 52 *thiS* construct predominantly exists in the aptamer configuration, which is distinct from even modestly longer forms that partially (53 *thiS*) or exclusively (54 *thiS*) adopt the terminator configuration (**Figure 4-5**). If HMP-PP is present and docks in the aptamer when the nascent mRNA transcript is exposed up to nucleotide position 52, terminator stem formation is precluded and transcription proceeds to generate the full-length RNA. Forward movement of RNA polymerase by even a single nucleotide yields a construct that begins to commit to the terminator stem structure, and at this point transcription would be terminated regardless of the presence of HMP-PP.

The *thiS* motif was originally identified as a riboswitch candidate by using an updated computational pipeline (Stav et al., 2019) that focuses attention on relatively long, GC-rich IGRs. This method of riboswitch discovery was developed to improve our ability to uncover rare and/or small riboswitch candidates. Thousands of riboswitch classes are predicted to remain undiscovered in bacteria (Ames and Breaker, 2010; McCown et al., 2017), and many of these are likely to be exceedingly rare and/or difficult to find because of their structural simplicity. Experimental demonstration that *thiS* motif RNAs are representatives of a compact riboswitch class that responds to HMP-PP provides initial confirmation that the updated computational pipeline can uncover riboswitch classes that have resisted discovery by other bioinformatic or genetic methods. We are hopeful that

many more riboswitch classes will be revealed by employing such bioinformatics search algorithms on additional sequenced bacterial genomes.

Finally, we note that the discovery of an HMP-PP riboswitch provides yet another example of a modern RNA class that selectively binds a molecule that was likely present in an RNA World (Benner et al., 1989; Gilbert, 1986). Many of the common enzyme cofactors are derived from RNA nucleotides or their precursors, including TPP and its precursor HMP-PP, which is a characteristic previously used to support the hypothesis that these molecules predate proteins (White III, 1976). TPP and many other RNA-derived coenzymes might have participated in a complex metabolic state run entirely by RNA enzymes and receptors, and therefore many riboswitches for coenzymes likewise might be molecular descendants from RNAs that utilized these coenzymes long ago (Breaker, 2011b; McCown et al., 2017). If true, HMP-PP riboswitches provide another opportunity to learn how simple RNA sequences and structures could have selectively bound important nucleotide-derived metabolites in primitive organisms of the RNA World.

Material and Methods

Oligonucleotides, chemicals and bacterial strains. Chemicals were purchased from Sigma-Aldrich with the exception of 4-amino-5-hydroxymethyl-2-methylpyrimidine, also known as (4-amino-2-methylpyrimidin-5-yl) methanol, which was purchased from Enamine Ltd. The radiolabeled molecules [γ-^{32}P]-ATP and [α-^{32}P]-ATP were purchased from PerkinElmer. All enzymes were purchased from New England BioLabs unless otherwise specified. *B. subtilis* strain 168 (BGSC 1A1) and the corresponding mutant strains *ΔthiS* (BGSC 11680), *ΔthiE* (BGSC 38290) and *ΔthiD* (BGSC 11710) were obtained from the Bacillus Genetic Stock Center (BGSC) at The Ohio State University. Plasmids pDG1661 and pDG148 were also obtained from BGSC. Synthetic DNA oligonucleotides were purchased from Sigma-Aldrich or Integrated DNA Technologies. A list of oligonucleotides used in this study can be found in **Table 4-1**.

Bioinformatic analysis. Representative *thiS* motif RNAs used to create the consensus sequence and structure models (**Figure 4-1A**) were identified using Infernal 1.1 (Nawrocki and Eddy, 2013) from RefSeq version 80 and certain environmental microbial databases as described previously (Weinberg et al., 2017). A total of 400 unique and complete representatives were used to generate an updated consensus model relative to that published previously (Stav et al., 2019). The consensus sequence and structural models were derived by using the R2R algorithm (Weinberg and Breaker, 2011), which employs weighting. To prevent irrelevant differences between the aptamer and terminator configurations caused by weighting, the consensus model of the aptamer configuration

was used to annotate the consensus sequence for the terminator configuration for the overlapping region.

Riboswitch reporter assays. Riboswitch-reporter constructs, integrating riboswitch representatives from either *C. sp* Maddingley or *C. lundense* DSM 17049, were prepared as synthetic oligonucleotides, amplified by PCR and cloned into vector pDG1661 upstream of the *E. coli lacZ* gene as described previously (Sudarsan et al., 2003; Nelson et al., 2017). Transcription initiation of the constructs is driven by the *B. subtilis lysC* gene promoter. The resulting WT and mutant reporter constructs were integrated into the *amyE* locus of WT (1A1 strain 168 *Δtrp)* or thiamin biosynthetic knockout strains (Δ*thiS*, Δ*thiE or* Δ*thiD*) as indicated. The resulting transformed strains were verified as previously described (Sherlock et al., 2018).

The *thiD* gene construct was generated by amplifying this gene from *B. subtilis* genomic DNA by PCR and inserted into the StuI site of a modified pDG148 vector using ligation-independent cloning as described previously (Joseph et al., 2001). The *lacI* gene in this vector has been mutated so that the *thiD* gene is expected to give constitutive expression. The resulting protein expression vector was then transformed into *B. subtilis* strains containing WT or mutant riboswitch reporter constructs, as indicated for each experiment.

Riboswitch-reporter assays were performed by inoculating various *B. subtilis* strains into Lysogeny Broth (LB) with appropriate antibiotics and growing overnight at 37ºC. For liquid-culture reporter assays with thiamin biosynthetic knock-out strains, overnight cultures grown in LB were then diluted 1/20 into Spizizen glucose minimal

medium (GMM) (Anagnostopoulos and Spizizen, 1961**)** and grown overnight at 37ºC. The residual thiamin from LB is sufficient for growth in GMM over the duration of the assay. For riboswitch reporter experiments with ThiD-producing strains, bacteria were diluted directly into LB and grown overnight with or without supplementation with HMP. Liquid media (LB or GMM) was supplemented with X-gal (200 μg mL^{-1}) to allow visual detection of reporter gene expression. Similarly, reporter expression analysis using 4-methylumbelliferyl β-D-galactopyranoside was conducted as described previously (Nelson et al., 2015; Atilho et al., 2019) to establish fluorescence units.

Expression and purification of ThiD. An N-terminal 6xHis-tagged *thiD* gene from *Salmonella typhimurium* was cloned into a pETDuet vector, which was then transformed into *E. coli* strain BL21(DE3). Transformed cells were grown in Terrific Broth medium until the OD_{600} reached 0.8. The resulting culture was incubated overnight at 16ºC for protein expression that was induced by the addition of 0.3 mM isopropyl β-D-1-thiogalactopyranoside. Cells were pelleted, resuspended in Buffer A [50 mM Tris (pH 8 at 23ºC), 400 mM NaCl, 10 mM imidazole, 5% glycerol, 0.1 mM tris(2-carboxyethyl)phosphine (TCEP)], and lysed with a microfluidizer. The resulting lysate was clarified via centrifugation, applied to NiNTA resin, and washed with 10 column volumes of Buffer A. Tagged protein was eluted from the column with three column volumes of Buffer B [50 mM Tris (pH 8 at 23ºC), 400 mM NaCl, 400 mM imidazole, 5% glycerol, 0.1 mM TCEP] and applied to a size-exclusion column equilibrated in Buffer C [50 mM Tris (pH 8 at 23ºC), 150 mM NaCl, 0.1 mM TCEP]. 10% Glycerol (v/v, final concentration) was added to the protein sample before storage at -80°C.

Enzymatic preparation of HMP-PP. Pyrophosphorylation of HMP was carried out enzymatically as described previously (Hanes et al., 2007). Briefly, the reaction was initiated by adding 10 μL of *Salmonella typhimurium* HMP-P kinase (ThiD, 20 mg/mL) to a 90 μL reaction preparation to yield a final concentration of 5 mM HMP, 20 mM ATP, 50 mM Tris-HCl buffer (pH 7.5 at 23ºC), 2 mM TCEP, and 5 mM $MgCl_2$. This 100 μL reaction mixture was allowed to incubate at 23ºC overnight, at which time the protein was removed by ultrafiltration using an Amicon Ultra-0.5 centrifugal filter unit with a 3 kDa cutoff membrane. The HMP-PP generated in this manner was used without further purification, typically on the same day but no later than 2 weeks after production. HMP-PP stock solutions were stored at -20ºC. Control assays for transcription termination and in-line probing conducted without the addition of enzymatically prepared HMP-PP contained an equivalent amount of the enzymatic reaction preparation wherein the ThiD protein was excluded. These control assays therefore contain HMP, whereas test reactions contain HMP-PP.

Mass spectrum analysis of HMP-PP. Enzymatically prepared HMP-PP samples were sent to the MS & Proteomics Resource at Yale University for analysis. The presence of HMP-PP was confirmed by data generated using a Thermo Scientific LTQ Orbitrap ELITE mass spectrometer. Data was acquired and analyzed with Xcalibur (v2.1). Peaks were considered to have the same mass-to-charge ratio as HMP-PP if they were within 10 ppm of the calculated ratio.

In vitro transcription termination assays. The protocol used for single-round *in vitro* transcription assays was adapted from that described previously (Landick et al., 1996). DNA constructs were designed to include the promoter sequence of the *lysC* gene from *B. subtilis*, the riboswitch aptamer, and the expression platform of the *thiS* gene from *C.* sp Maddignly to 33 nucleotides following the terminator stem. Additional non-native nucleotides were added to the 5′ region upstream of the HMP-PP aptamer to increase the amount of [α-^{32}P]-ATP incorporation.

To assemble each *in vitro* transcription reaction, approximately 2 pmol of the purified, PCR amplified DNA template was added to a transcription initiation mixture [final concentration of 20 mM Tris (pH 8.0 at 23ºC), 20 mM NaCl, 14 mM $MgCl_2$, 100 μM EDTA, 10 μg mL^{-1} bovine serum albumin, 130 μM ApA dinucleotide, 1% glycerol, 0.04 U μL^{-1} *E. coli* RNA polymerase holoenzyme, 2.5 μM GTP, 2.5 μM UTP, and 1 μM ATP]. Approximately 8 μCi [α-^{32}P]-ATP was added to the 90 μL transcription reaction and transcription was allowed to proceed at 37ºC for 30 minutes, leading to formation of a stalled polymerase complex before the first cytidine of each transcript. The reaction mixture was then distributed in 8 μL aliquots into separate microfuge tubes, which contained 1 μL of an HMP or HMP-PP solution plus 1 μL of 10x elongation buffer (200 mM Tris [pH 8.0 at 23ºC], 200 mM NaCl, 140 mM $MgCl_2$, 1 mM EDTA, 1 mg mL^{-1} heparin, 1.5 mM each of ATP, GTP, and CTP, and 0.5 mM UTP). Transcription elongation was allowed to proceed for 45 minutes at 37ºC.

The transcription products were separated by denaturing (8 M urea) 10% polyacrylamide gel electrophoresis (PAGE) then imaged and quantified using a Typhoon Phosphorimager and ImageQuaNT software. The fraction of full length (FL) and

terminated (T) RNA transcripts was calculated by measuring band intensity values and using the equation Fraction FL = (FL intensity)/(FL intensity + T intensity). The differences in specific activities between the FL and T products due to [α-^{32}P]-ATP incorporation were considered negligible.

RNA oligonucleotide preparation. RNAs were prepared by *in vitro* transcription using DNA oligonucleotides containing a T7 RNA polymerase promoter sequence upstream of the desired template sequence. The resulting desired RNA transcripts were purified, enzymatically 5′ ^{32}P-labeled, and repurified as previously described (Arachchilage et al., 2018; Atilho et al., 2019).

RNA in-line probing analysis. In-line probing assays (Soukup and Breaker, 1999; Regulski and Breaker, 2008) were performed precisely as described previously (Atilho et al., 2019; Mirihana Arachchilage et al., 2018). The maximum HMP-PP concentration (max) is achieved by using a 3/10 dilution of enzymatically prepared HMP-PP as described above. The control reaction was performed using the 1/10 dilution of the control sample that lacks HMP-PP.

References

Ames TD, Breaker RR. 2010. Bacterial riboswitch discovery and analysis. In *The Chemical Biology of Nucleic Acids* (ed. Mayer G). Wiley, Chichester, UK.

Anagnostopoulos C, Spizizen J. 1961. Requirements for transformation in *Bacillus subtilis*. *Journal of Bacteriology* **81**: 741–746.

Atilho RM, Perkins KR, Breaker RR. 2019. Rare variants of the FMN riboswitch class in *Clostridium difficile* and other bacteria exhibit altered ligand specificity. *RNA* **25**: 23-34.

Barrick JE, Breaker RR. 2007. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biology* **8**:R239.

Begley TP, Ealick SE, McLafferty FW. 2012. Thiamin biosynthesis – still yielding fascinating biological chemistry. *Biochemical Society Transactions* **40:** 555-560.

Benner SA, Ellington AD, Tauer A. 1989. Modern metabolism as a palimpsest of the RNA world. *Proceedings of the National Academy of Sciences USA* **86**: 7054-7058.

Breaker RR. 2011a. Prospects for riboswitch discovery and analysis. *Molecular Cell* **43:** 867-879.

Breaker RR. 2011b. Riboswitches and the RNA World. *Cold Spring Harbor Perspectives in Biology* 4, a003566.

Gilbert W. 1986. Origin of life: The RNA world. *Nature* **319:** 618.

Hanes JW, Ealick SE, Begley TP. 2007. Thiamin phosphate synthase: the rate of pyrimidine carbocation formation. *Journal of the American Chemical Society* **129:** 4860-4861.

Hein PP, Kolb KE, Windgassen T, Bellecourt MJ, Darst SA, Mooney RA, Landick R. 2014. RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nature Structural Biology* **21**: 794-802.

Joseph P, Fantino JR, Herbaud ML, Denizot F. 2001. Rapid orientated cloning in a shuttle vector allowing modulated gene expression in *Bacillus subtilis*. *FEMS Microbiology Letters* **205**: 91-97.

Jurgenson CT, Begley TP, Ealick SE. 2009. The structural and biochemical foundations of thiamin biosynthesis. *Annual Review of Biochemistry* **78:** 569-603.

Klein RJ, Misulovin Z, Eddy SR. 2002. Noncoding RNA genes identified in AT-rich hyperthermophiles. *Proceedings of the National Academy of Sciences* **99:** 7542-7547.

Komissarova N, Kashlev M. 1998. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proceedings of the National Academy of Sciences* **95:** 14699-14704.

Landick R, Wang D, Chan CL. 1996. Quantitative analysis of transcriptional pausing by *Escherichia coli* RNA polymerase: his leader pause site as paradigm. *Methods in Enzymology* **274**: 334-353.

Lee ER, Baker JL, Weinberg Z, Sudarsan N, Breaker RR. 2010. An allosteric self-splicing ribozyme triggered by a bacterial second messenger. *Science* **329:** 845-848.

McCown PJ, Corbino KA, Stav S, Sherlock ME, Breaker RR. 2017. Riboswitch diversity and distribution. *RNA* **23:** 995-1011.

McDaniel BA, Grundy FJ, Artsimovitch I, Henkin TM. 2003. Transcription termination control of the S box system: direct measurement of *S*-adenosylmethionine by the leader RNA. *Proceedings of the National Academy of Sciences* **100:** 3083-3088.

Meyer MM, Ames TD, Smith DP, Weinberg Z, Schwalbach MS, Giovanni SJ, Breaker RR. 2009. Identification of candidate structured RNAs in the marine organism '*Candidatus Pelagibacter ubique*'. *BMC Genomics* **10:** 268.

Mironov AS, Gusarov I, Rafikov R, Lopez LE, Shatalin K, Kreneva RA, Perumov DA, Nudler E. 2002. Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* 111: 747-56.

Mirihana Arachchilage G, Sherlock ME, Weinberg Z, Breaker RR. 2018. SAM-VI RNAs selectively bind *S*-adenosylmethionine and exhibit similarities to SAM-III riboswitches. *RNA Biology* **15**: 371-378.

Monforte JA, Kahn JD, Hearst JE. 1990. RNA folding during transcription by *Escherichia coli* polymerase analyzed by RNA self-cleavage. *Biochemistry* **29:** 7882-7890.

Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29:** 2933-2935.

Nelson JW, Plummer MS, Blount KF, Ames TD, Breaker RR. 2015 Small molecule fluoride toxicity agonists. *Chemical Biology* **22:** 527-534.

Nelson JW, Atilho RM, Sherlock ME, Stockbridge RB, Breaker RR. 2017. Metabolism of free guanidine in bacteria is regulated by a widespread riboswitch class. *Molecular Cell* **65:** 220-230.

Nelson JW, Breaker RR. 2017. The lost language of the RNA World. *Science Signaling* **10:** 1-10.

Poiata E, Meyer MM, Ames TD, Breaker RR. 2009. A variant riboswitch aptamer class for *S*-adenosylmethionine common in marine bacteria. *RNA* **15:** 2046-2056.

Regulski EE, Breaker RR. 2008. In-line probing analysis of riboswitches. *Methods Molecular Biology* **419:** 53-67.

Rosewarne CP, Greenfield P, Li D, Tran-Dinh N, Bradbury MI, Midgley DJ, Hendry P. 2013. Draft genome sequence of *Clostridium* sp. Maddingley, isolated from coal-seam gas formation water. *Genome Announcements* **1:**pii: e00081-12.

Schattner P. 2002. Searching for RNA genes using base-composition statistics. *Nucleic Acids Research* **30:** 2076-2082.

Serganov A, Nudler E. 2013. A decade of riboswitches. *Cell* **152**: 17-24.

Sherlock ME, Sudarsan N, Stav S, Breaker RR. 2018. Tandem riboswitches form a natural Boolean logic gate to control purine metabolism in bacteria. *eLife* **7:** e33908.

Sherwood AV, Henkin TM. 2016. Riboswitch-mediated gene regulation: Novel RNA architectures dictate gene expression responses. *Annual Review of Microbiology* **70:** 361-374.

Soukup GA, Breaker RR. 1999. Relationship between internucleotide linkage geometry and the stability of RNA. *RNA* **5:** 1308-1325.

Stav S, Atilho RM, Mirihana Arachchilage G, Nguyen G, Higgs G, Breaker RR. 2019. Genome-wide discovery of structured noncoding RNAs in bacteria. (submitted)

Stoddard CD, Batey RT. 2006. Mix and match riboswitches. *ACS Chemical Biology* **1**: 751-754.

Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. 2003. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes and Development* **17:** 2688-2697.

Sudarsan N, Hammond MC, Block KF, Welz R, Barrick JE, Roth A, Breaker RR. 2006. Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314**: 300-304.

Weinberg Z, Nelson JW, Lünse CE, Sherlock ME, Breaker RR. 2017. Bioinformatic analysis of riboswitch structures uncovers variant classes with altered ligand specificity. *Proceedings of the National Academy of Sciences USA* **114:** 2077–2085.

Weinberg Z, Breaker RR. 2011. R2R - software to speed the depiction of aesthetic consensus RNA secondary structures. *BMC Bioinformatics* **12:** 3.

White III HB. 1976. Coenzymes as fossils of an earlier metabolic state. *Journal of Molecular Evolution* **7:** 101-104.

Wickiser JK, Winkler WC, Breaker RR, Crothers DM. 2005a. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Molecular Cell* **18:** 49-60.

Wickiser JK, Cheah MT, Breaker RR, Crothers DM. 2005b. The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* **44:** 13404-13414.

Wilson KS, von Hippel PH. 1995. Transcription termination at intrinsic terminators: the role of the RNA hairpin. *Proceedings of the National Academy of Sciences* **92:** 8793-8797.

Yarnell WS, Roberts JW. 1999. Mechanism of intrinsic transcription termination and antitermination. *Science* **284**: 611-615.

Figures and Tables

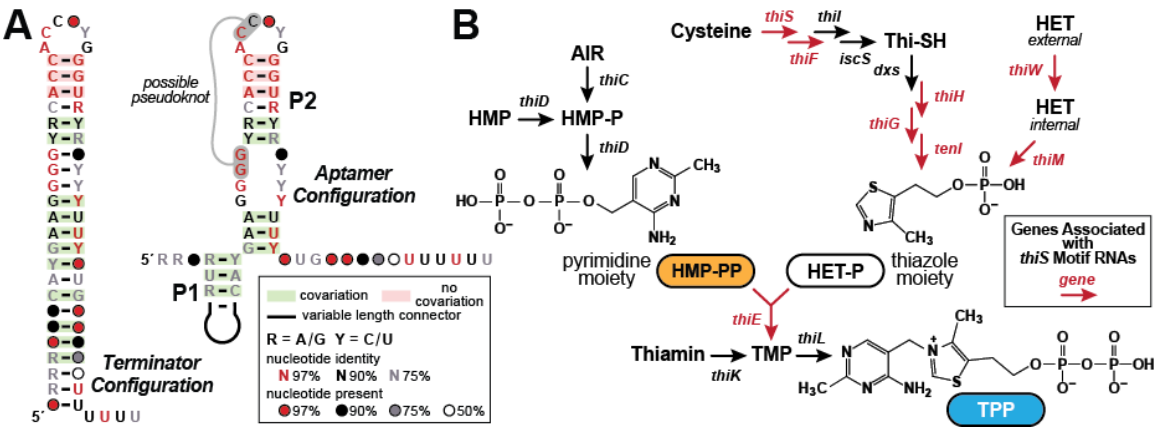

Figure 4-1. The *thiS* motif consensus models and gene associations. (A) Conserved nucleotide sequences and two possible secondary structure models based on 400 representatives of *thiS* motif RNAs. (Left) Terminator configuration exhibits features (strong stem followed by a run of U nucleotides) that are consistent with a typical bacterial intrinsic transcription terminator stem ('OFF' state). (Right) Another secondary structure model ('ON' state) for *thiS* motif RNAs, consistent with the pattern of conserved base-pairing potential, involves the formation of distinct base-pairing to form a P1 stem and a possible pseudoknot, whereas the base-pairing of P2 is identical to the tip of the terminator stem configuration. (B) Typical biosynthetic pathways for the pyrimidine and thiazole moieties of thiamin, and their eventual conversion into the enzyme cofactor thiamin pyrophosphate (TPP). Genes coding for the protein enzymes involved in HET transport and catalysis of various biosynthetic steps are annotated with colors that reflect their association with *thiS* motif representatives.

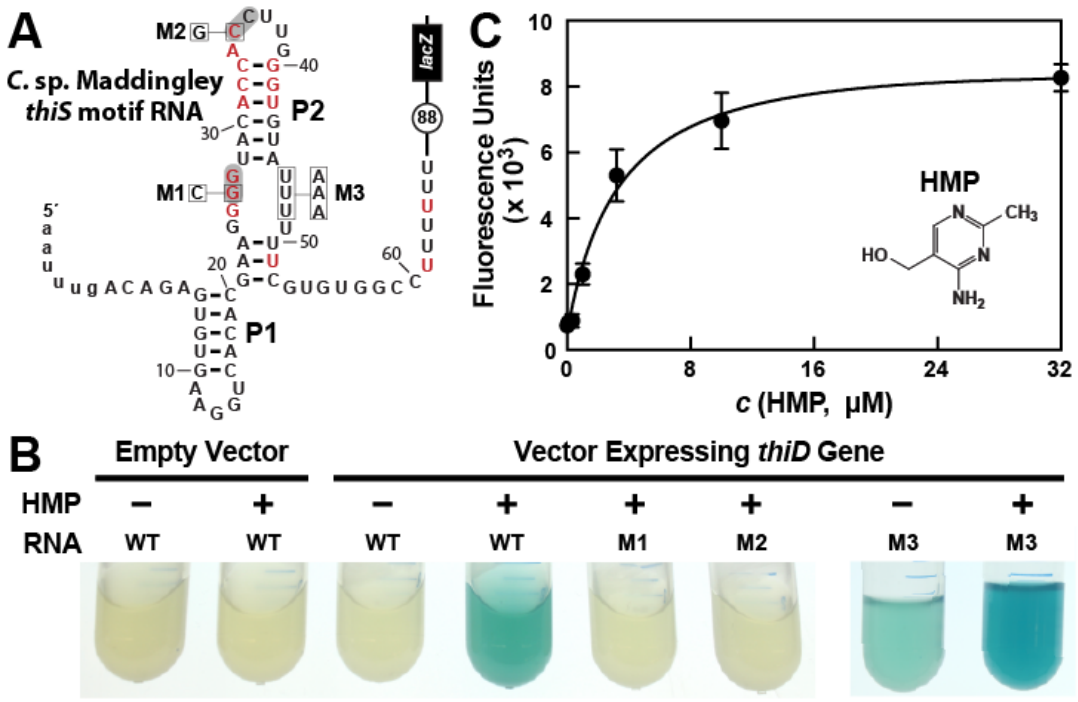

Figure 4-2. HMP-PP triggers reporter gene expression. (A) Sequence and predicted secondary structure of the WT *thiS* motif RNA associated with the *thiS* gene of *C.* sp. Maddingley fused to a β-galactosidase reporter gene (*lacZ*). The structural model is depicted in its presumed 'ON' state, where the shaded nucleotides highlight possible base-pairs of the predicted pseudoknot. The encircled number represents the additional nucleotides linking the end of the intrinsic terminator element (6 U nucleotides) and the coding region joined to the *lacZ* reporter gene sequence. Red nucleotides are >97% conserved as depicted in the *thiS* consensus model (**Figure 4-1A**). The nucleotides are numbered beginning with the predicted natural transcription start site, and the lowercase letters at the 5′ end identify additional nucleotides included in the reporter-fusion construct. Boxed nucleotides identify positions that are altered in the mutant constructs indicated. (B) Photos of reporter gene assay liquid cultures of *B. subtilis* cells carrying the protein expression plasmid pDG148-Stu either lacking (empty vector) or carrying a

native *thiD* gene. Reporter constructs carrying the WT riboswitch construct or mutations M1 through M3 as depicted in A were grown in LB either without (–) or with HMP supplementation. (C) HMP concentration dependent reporter gene expression of WT *B. subtilis* cells carrying the riboswitch-*lacZ* fusion construct depicted in A. Cells were grown in minimal (GMM) liquid medium supplemented with various concentrations of HMP as indicated. Data points are the average of three measurements and are representative of experiments performed on multiple days. Error bars indicate the standard deviation of the measurements, which when not visible are smaller than the corresponding data points.

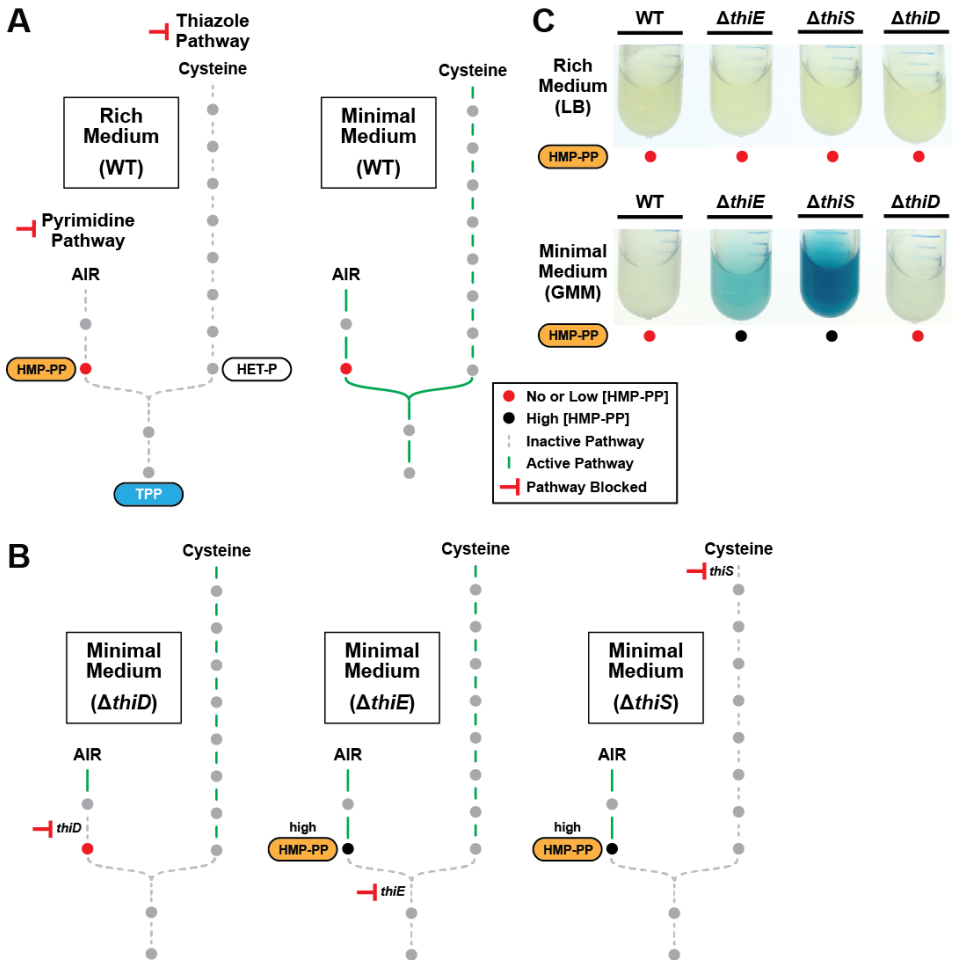

Figure 4-3. The effects of thiamin biosynthetic pathway knockouts on expression of a *thiS* riboswitch-reporter construct. (A) Schematic representation of the expected metabolic activities of the pyrimidine and thiazole pathways for TPP biosynthesis in rich (LB, left) and minimal (GMM, right) media. Solid green lines indicate pathways with flux, whereas dashed gray lines represent pathways with low or no flux. Shaded circles represent molecules in the pathway. (B) Schematic representation of the expected metabolic flux of thiamin knockout strains grown in minimal medium. Other annotations are as described for panel A. (C) Reporter gene expression of WT *B. subtilis* cells and cells lacking the coding region for various thiamin biosynthesis genes grown in rich or minimal liquid media.

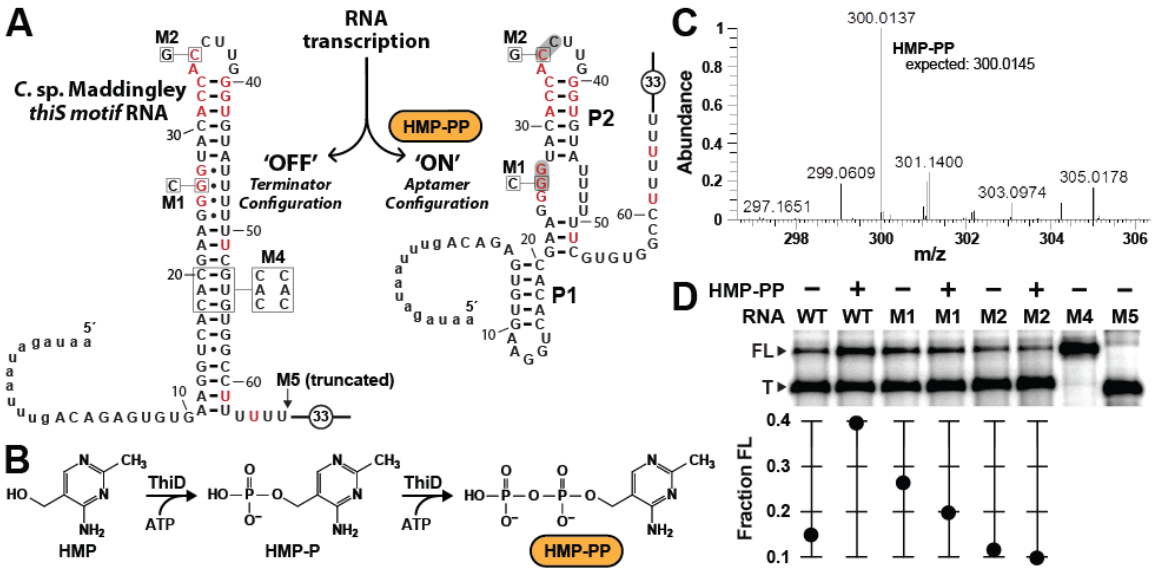

Figure 4-4. Pyrophosphorylated HMP triggers riboswitch-mediated transcription elongation in vitro. (A) Sequence and secondary structure models for the 'OFF' and 'ON' states of an HMP-PP riboswitch construct from *C.* sp. Maddingley used for transcription termination assays. In the absence of ligand ('OFF' state, left), an intrinsic transcription terminator stem is expected to form and cause transcription termination near nucleotide number 65. In the ligand-bound structure ('ON' state, right), a distinct P1 stem and a P2 stem (partial formation of the terminator) are expected to form, thereby blocking formation of the complete terminator stem and leading to transcription read-through. (B) Pathway for the enzymatic pyrophosphorylation of HMP using 4-amino-5-hydroxymethyl-2-methylpyrimidine phosphate kinase (ThiD), carried out as previously described (Hanes et al., 2007). (**C**) Plots of the mass-to-charge ratios (m/z) of HMP and HMP-PP prepared by enzymatic pyrophosphorylation. Relative abundance (peak heights) normalized to that of HMP-PP. (D) (Top) PAGE analysis of single-round transcription termination assays of WT and various mutant HMP-PP riboswitch constructs in the

absence or presence of ligand. ‘FL’ and ‘T’ denote the full length and terminated transcripts, respectively. (Bottom) Values for the fraction of full-length RNA transcript relative to the total transcription yield are plotted for each reaction.

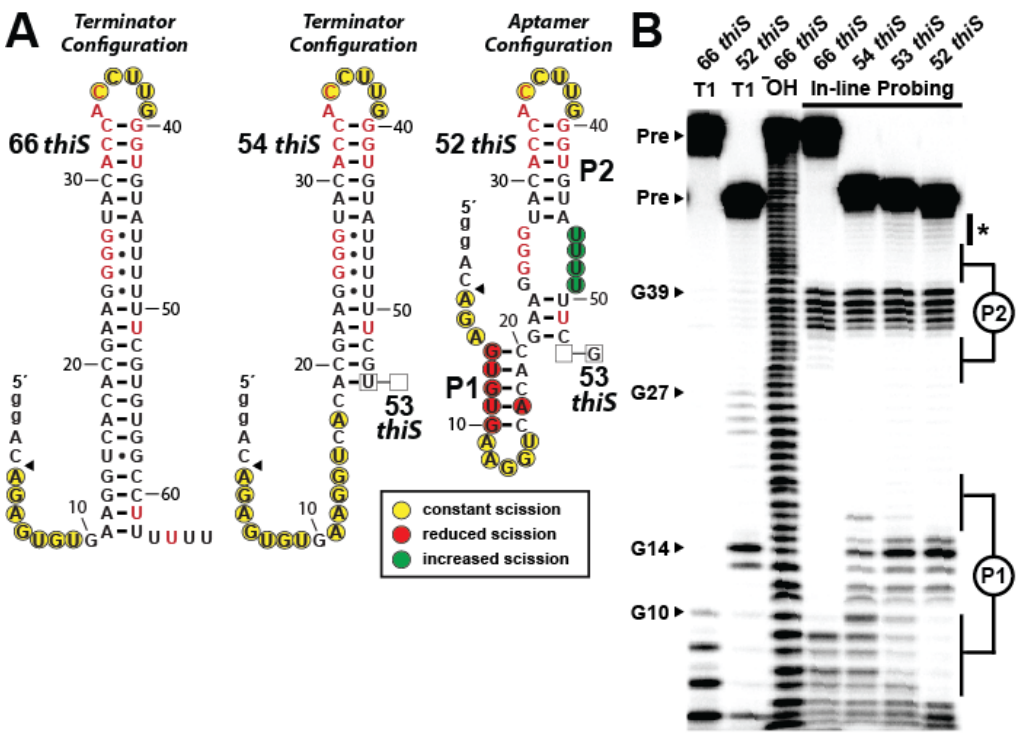

Figure 4-5. Substantial conformational rearrangement of an HMP-PP riboswitch is brought about by single-nucleotide differences at the 3´ terminus. (A) Sequences and secondary structure models of RNA constructs used to assess conformational determination between terminator and aptamer structures. Nucleotide differences between constructs are boxed, and empty boxes indicates the nucleotide is absent. The secondary structure models were predicted based on the in-line probing data derived from B. Note that the 53 *thiS* RNA secondary structure model adopts both the terminator (predominant form for 54 *thiS*) and aptamer (predominant form for 52 *thiS*) conformations with near equal probability. Yellow circles identify nucleotide linkages that undergo robust spontaneous strand scission regardless of the length of the construct, suggesting they always remain unstructured. Red circles identify linkages that exhibit reduced spontaneous scission compared to larger constructs, suggesting the corresponding nucleotides become more structured in the 52 *thiS* construct. Green circles

identify linkages undergoing increased strand scission, suggesting these nucleotides become less structured in 52 *thiS* construct. The arrowhead identifies the start of the interpretable data in B. (B) PAGE analysis of 5′ ^{32}P-radiolabeled RNA constructs subjected to in-line probing analysis in the absence of HMP-PP. T1 and $^{-}$OH indicate RNase T1 partial digestion (cleavage after G nucleotides) and alkaline-mediated partial digestion (cleaves after each nucleotide), respectively. Bands corresponding to precursor RNA (Pre) and enzymatic cleavage after certain G residues are annotated. The asterisk identifies bands corresponding to nucleotides 46-49 of 52 *thiS*, which appear to increase in intensity compared to the 66 *thiS* construct.

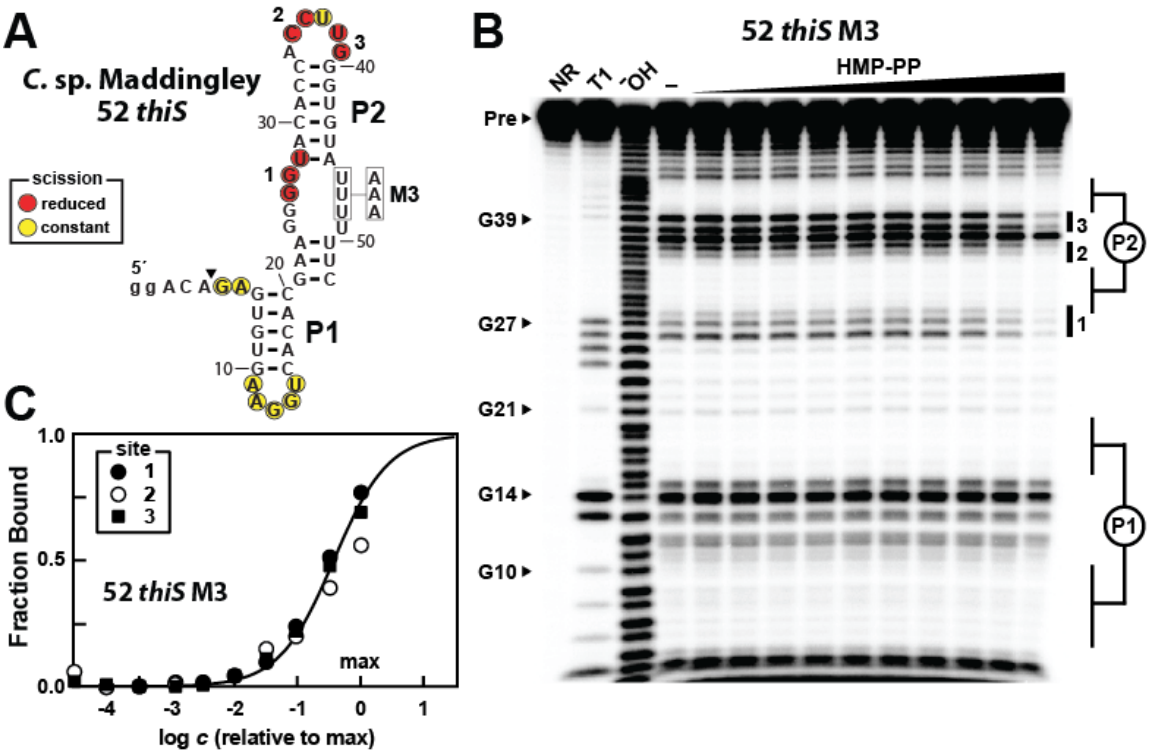

Figure 4-6. A modified representative HMP-PP riboswitch aptamer binds HMP-PP. (A) Sequence and secondary structure model of the 52 *thiS* RNA construct derived from the *thiS* gene of *C.* sp. Maddingley, and its M3 mutant (U46A, U47A, U48A). Both the WT and M3 constructs are known to retain gene control activity in response to ligand (**Figure 4-2**). Nucleotides at the 5′ end denoted in lowercase letters were added to enhance the efficiency of transcription *in vitro*. The characteristics of sites of spontaneous RNA cleavage resulting from in-line probing reactions in the absence or presence of HMP-PP, as derived from the image in B, are indicated. (B) PAGE analysis of 5′ ^{32}P-radiolabeled 52 *thiS* M3 RNA that was subjected to in-line probing assays in the absence (–) or presence of HMP-PP. The ligand HMP-PP was enzymatically produced using ThiD protein and used in a range of concentrations relative to the maximum HMP-PP concentration (precise concentration unknown, see Materials and Methods). NR, T1 and $^{-}$OH indicate no reaction, RNase T1 partial digestion (cleavage after G nucleotides), and

alkaline-mediated partial digestion (cleaves after each nucleotide), respectively. Bands corresponding to precursor RNA (Pre) and enzymatic cleavage after certain G residues are annotated. Notable sites of ligand-mediated structural modulation are numbered 1 through 3. (C) Plot depicting the fraction of RNA bound to ligand versus the logarithm of the concentration of HMP-PP relative to the maximum (max) ligand concentration tested. The data points were generated by quantifying band intensity changes at sites 1, 2 and 3 in B with correction for loading and yield differences, including the general reduction in RNA cleavage observed in the final lane containing the maximum HMP-PP concentration tested. The line presents a theoretical 1:1 binding curve (Hill coefficient of 1) and is overlaid on the data points for comparison.

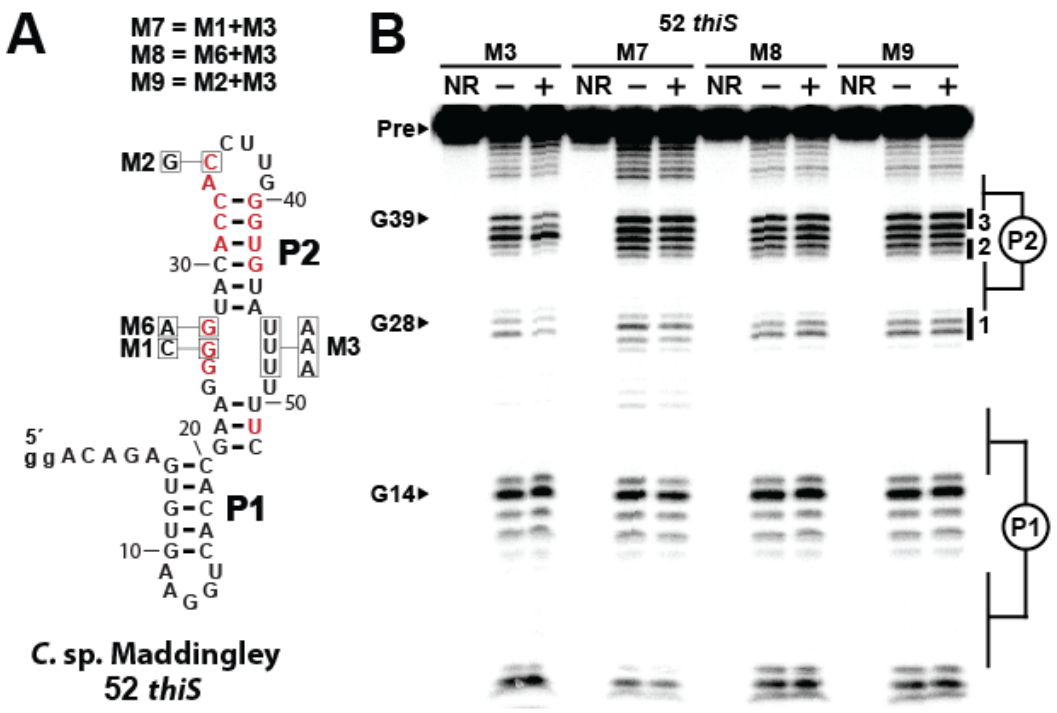

Figure 4-7. Mutational analysis of an HMP-PP aptamer. (A) Sequence and secondary structure model of the 52 *thiS* parent sequence and the mutations used to form the variant constructs. Nucleotides in red identify the highly conserved (>97%) positions as determined by comparative sequence analysis (**Figure 4-1A**). Nucleotide positions that were mutated in each RNA construct are boxed and annotated with the mutant nucleotide identity. The mutant constructs M6 through M8 were obtained by including M1, M6, and M2 mutations in the 52 *thiS* M3 RNA construct, which originally exhibits direct binding of HMP-PP (**Figure 4-6**). (B) PAGE analysis of in-line probing assays performed with 52 *thiS* M3 RNA, or with each mutant RNA construct (M7 through M9) as indicated in the absence (–) or presence (+) of HMP-PP. Additional annotations are as described in the legend to **Figure 4-6**.

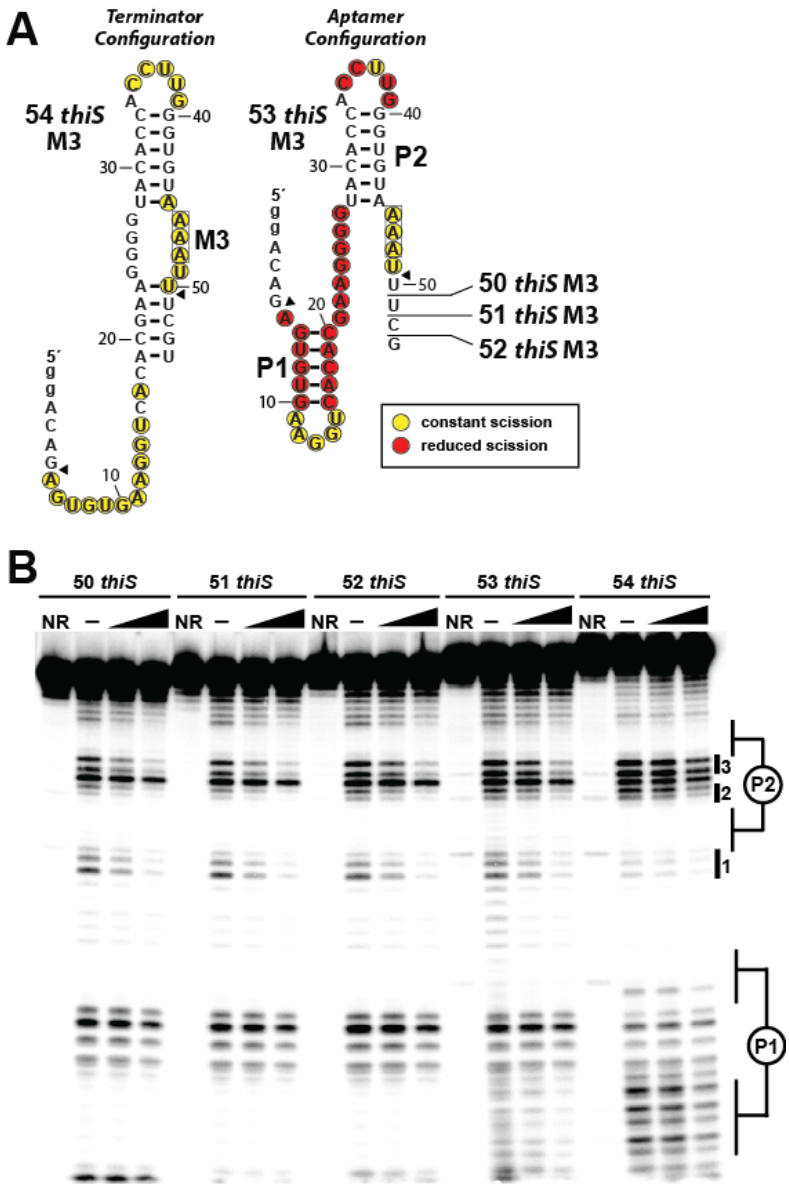

Figure 4-8. HMP-PP mediated conformational rearrangement of construct 53 thiS M3 and related RNAs. (A) Sequences and secondary structure models for five progressively shorter *thiS* M3 RNA constructs. The 54 *thiS* M3 construct (left) is identical to the 54 *thiS* construct depicted in the main text (**Figure 4-5A**), except that three U nucleotides spanning positions 46-48 have been converted to A nucleotides (M3). The remaining constructs are each, consecutively, shortened one nucleotide at their 3′ termini, to yield constructs 53 *thiS* M3 through 50 *thiS* M3. Scission annotations for 54

thiS M3 are derived from the data presented in B in the absence of HMP-PP. Scission annotations for 53 *thiS* M3 are derived from the data presented in B, and reflect the difference in banding pattern between the (–) and max HMP-PP lanes. (B) PAGE analysis of 5′ ^{32}P-radiolabeled M3 mutants of 50 *thiS* through 54 *thiS* RNA constructs, subjected to in-line probing analysis in absence or presence of HMP-PP. The two concentrations of ligand used are the 1 and 1/3 of the maximum concentration (max) of HMP-PP as described in the main text (see legend to **Figure 4-6**). NR indicates no reaction and (–) indicates no HMP-PP added. Regions of the banding pattern corresponding to the locations of stems (P1 and P2) and notable sites of modulation (1 through 3) are denoted as observed previously (**Figure 4-5B**, **Figure 4-6B**).

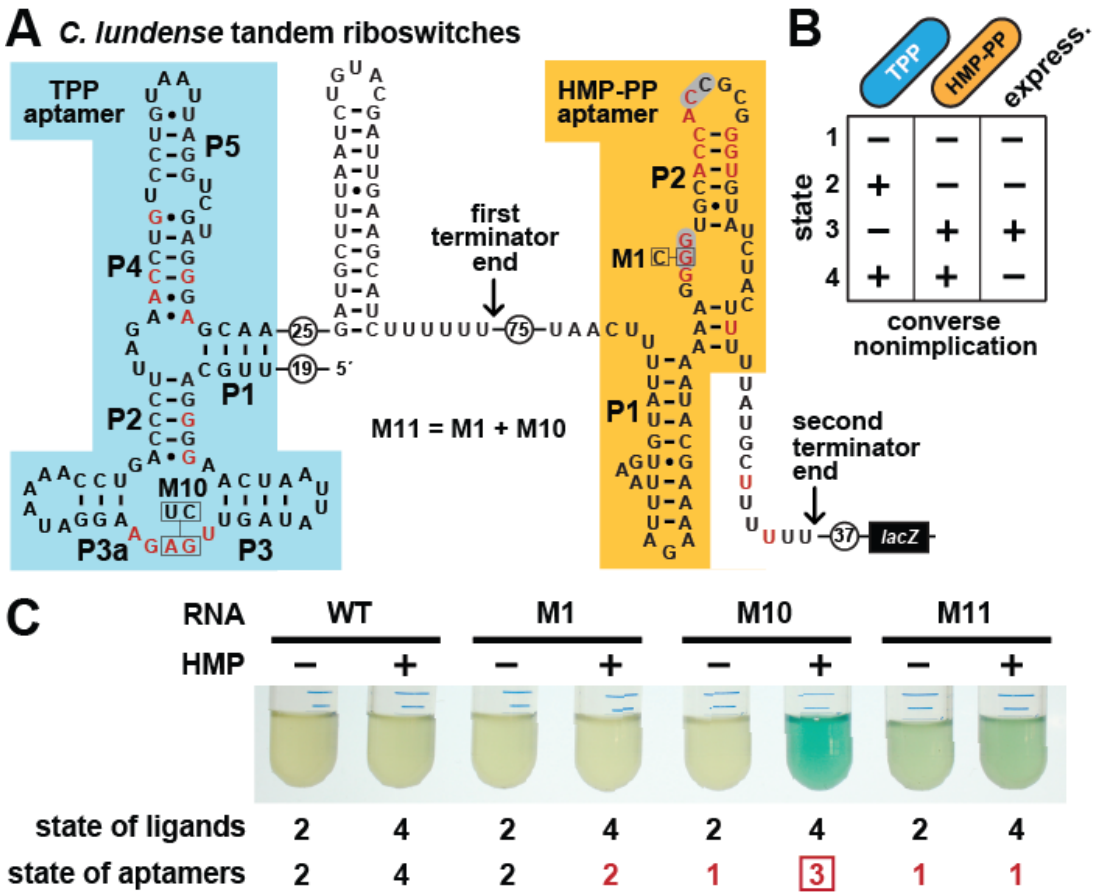

Figure 4-9. Tandem TPP and HMP-PP riboswitches form a two-input Boolean logic gate to regulate gene expression. (A) Sequence and secondary structure of a tandem riboswitch arrangement derived from the *thiW* gene from *C. lundense* DSM 17049 that is responsive to TPP and HMP-PP ligands. Red letters identify positions matching the most highly conserved nucleotides in the consensus sequence models for TPP (McCown et al., 2017) and HMP-PP (**Figure 4-1A**) aptamers. (B) Predicted truth table for tandem TPP and HMP-PP riboswitches represents a converse nonimplication logic gate. Minus and plus symbols, respectively, represent low or high gene expression (express.) or ligand concentrations as indicated. (C) Reporter gene expression of a *B. subtilis* strain expressing ThiD, and carrying either a wild-type (WT) or a mutant (M1, M10 or M11) *C. lundense* tandem riboswitch-*lacZ* reporter fusion construct as described in A. Cells were

grown in rich (LB) medium containing either no (–) or 10 μM additional HMP and 200 μg mL^{-1} X-gal. The numbers listed as 'state of ligands' represents the cellular presence or absence of ligands corresponding to the 'states' depicted in the truth table in B. The numbers listed as 'state of aptamers' represents the expected status (free or ligand bound) of the aptamers in the riboswitch, again as depicted in the truth table. Red numbers identify states artificially created by riboswitch mutations. The box identifies state 3, which is created by the addition of HMP to construct M10.

Table 4-1. Sequences of synthetic DNAs used in Chapter 4.

Name	Sequence (5′ to 3′)	Annotation
thiS1	AAAAAAGGCCACACGAAAAAATACACCCAAGGTGGTGTACCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of *C.* sp. Maddingley 66 *thiS* RNA
thiS2	ACGAAAAAATACACCCAAGGTGGTGTACCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of *C.* sp. Maddingley 54 *thiS* RNA
thiS3	CGAAAAAATACACCCAAGGTGGTGTACCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of *C.* sp. Maddingley 53 *thiS* RNA
thiS4	GAAAAAATACACCCAAGGTGGTGTACCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of *C.* sp. Maddingley 52 *thiS* RNA
thiS5	AATTTTACACCCAAGGTGGTGTACCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of M3 50 *thiS* RNA RNA
thiS6	AAATTTTACACCCAAGGTGGTGTACCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of M3 51 *thiS* RNA RNA
thiS7	GAAATTTTACACCCAAGGTGGTGTACCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of M3 52 *thiS* RNA RNA
thiS8	CGAAATTTTACACCCAAGGTGGTGTACCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTA	The template for transcription of M3 53 *thiS* RNA RNA

	TTA	
thiS9	ACGAAATTTTACACCCAAGGTGGTGTACCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of M3 54 *thiS* RNA RNA
thiS10	GAAATTTTACACCCAAGGTGGTGTACGCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of M7 RNA
thiS11	GAAATTTTACACCCAAGGTGGTGTATCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of M8 RNA
thiS12	GAAATTTTACACCCAAGCTGGTGTACCCCTTCGTGTGACCTTCACACTCTGTCCTATAGTGAGTCGTATTA	The template for transcription of M9 RNA
thiS13	TAATACGACTCACTATA	The sense strand of T7 promoter used for *in vitro* transcription
thiS14	TACGACGAATTCCAAAAATAATGTTGATCCTTTTAAATAAGTCTGATAAAATGTGAACTAATTTGACAGAGTGTGAAGGTCACACGAAGGGGTACACCACCTTGGGTGTATTTTTTCGTGTGGCCTTTTTTGCATGTATAAATAGGAAGAGGTGAAAAGTGGATCCAAAGGA	IDT G-block containing the wild-type *C.* sp. Maddingley *thiS* riboswitch and controlled by the *B. subtilis lysC* promoter, for cloning into pDG1661 as well as serving as a template for the dsDNA template for *in vitro* transcription termination
thiS15	TACGACGAATTCCAAAAATAATGTTGATCCTTTTAAATAAGTC	Forward primer for the PCR amplification of the *C.* sp. Maddingley *thiS* reporter construct
thiS16	TCCTTTGGATCCACTTTTCACCTCTTCC	Reverse primer for the PCR amplification of the *C.* sp. Maddingley *thiS* reporter construct

thiS17	GTGAAGGTCACACGAAGGCGTA CACCACCTTGGGTG	Forward primer for construction of the *C.* sp. Maddingley *thiS* reporter as well as single-round transcription termination template containing M1 via QuikChange mutagenesis
thiS18	CACCCAAGGTGGTGTACGCCTTC GTGTGACCTTCAC	Reverse primer for construction of the *C.* sp. Maddingley *thiS* reporter as well as single-round transcription termination template containing M1 via QuikChange mutagenesis
thiS19	CACACGAAGGGGTACACCAGCT TGGGTGTATTTTTTCG	Forward primer for construction of the *C.* sp. Maddingley *thiS* reporter as well as single-round transcription termination template containing M2 via QuikChange mutagenesis
thiS20	CGAAAAAATACACCCAAGCTGG TGTACCCCTTCGTGTG	Reverse primer for construction of the *C.* sp. Maddingley *thiS* reporter as well as single-round transcription termination template containing M2 via QuikChange mutagenesis
thiS21	GTACACCACCTTGGGTGTAAAAT TTCGTGTGGCCTTTTTTG	Forward primer for construction of the *C.* sp. Maddingley *thiS* reporter as well as single-round transcription termination template containing M3 via QuikChange mutagenesis
thiS22	CAAAAAAGGCCACACGAAATTT TACACCCAAGGTGGTGTAC	Reverse primer for construction of the *C.* sp. Maddingley *thiS* reporter as well as single-round transcription termination template containing M3 via QuikChange mutagenesis
thiS23	CTTGGGTGTATTTTTTCCACTGGC CTTTTTTGCATG	Forward primer for construction of the *C.* sp. Maddingley *thiS* single-round transcription termination template containing M4 via QuikChange mutagenesis
thiS24	CATGCAAAAAAGGCCAGTGGAA	Reverse primer for construction of the *C.*

	AAAATACACCCAAG	sp. Maddingley *thiS* reporter as well as single-round transcription termination template containing M4 via QuikChange mutagenesis
thiS25	CAAAAATAATGTTGATCCTTTTA AATAAGTCTGATAAAATGTGAAC TAATAGATATTTGACAGAGTGTG AAGGTC	Forward primer for the PCR amplification of the *C.* sp. Maddingley *thiS* IVT template
thiS26	ATCCACTTTTCACCTCTTCCTATT TATAC	Reverse primer for the PCR amplification of the *C.* sp. Maddingley *thiS* IVT template
thiS27	AAAAAAGGCCACACGAAAAAAT ACACCC	Reverse primer for the PCR amplification of the *C.* sp. Maddingley *thiS* M5 IVT template
thiS28	TACGACGAATTCCAACAGTAATG TTGATCCTTTTAAATAAGTCTGA TAAAATGTGAACTAAATAAAATT TGATATAAATTGCAGGGGAACTA ATTATAGTTGAGAAGGATAAAA CCTGACCCTTAGAACCTGTCAGT TAATACTGTCGTAGGGAGCAAA GCACTGTGATTAAGATAAGATAA AGATGCTTTAATCTGTACGATTG AAGCATCTTTTTTAAATTTAATA TTAATTAATCAGATTTTCAGTTT AAAGTAGGAAAATTTGAAACAG AAAACTTAATAAGTAACACGTA ACTTTTATGTGAATTTAGAAAAG CATAAAAAGGGGTGCACCACCG CGGGTGTATCTACTTTTTATGCTT TTTTATTTAAAATTTAATAGATC AGACTTGGATCCAGCTGCA	IDT G-block containing the *C. lundense* tandem reporter controlled by the *B. subtilis lysC* promoter for cloning into pDG1661.
thiS29	TACGACGAATTCCAACAGTAATG TTGATCCTTTTAAATAAGTCTGA TAAAATGTGAACTAAATAAAATT TGATATAAATTGCAGGGGAACTA	IDT G-block containing the M1 *C. lundense* tandem reporter controlled by the *B. subtilis lysC* promoter for cloning into pDG1661.

	ATTATAGTTGAGAAGGATAAAA CCTGACCCTTAGAACCTGTCAGT TAATACTGTCGTAGGGAGCAAA GCACTGTGATTAAGATAAGATAA AGATGCTTTAATCTGTACGATTG AAGCATCTTTTTTAAATTTAATA TTAATTAATCAGATTTTCAGTTT AAAGTAGGAAAATTTGAAACAG AAAACTTAATAAGTAACACGTA ACTTTTATGTGAATTTAGAAAAG CATAAAAAGGCGTGCACCACCG CGGGTGTATCTACTTTTTATGCTT TTTTATTTAAAATTTAATAGATC AGACTTGGATCCAGCTGCA	
thiS30	TACGACGAATTCCAACAGTAATG TTGATCCTTTTAAATAAGTCTGA TAAAATGTGAACTAAATAAAATT TGATATAAATTGCAGGGGAACTA ATTATAGTTCTGAAGGATAAAAC CTGACCCTTAGAACCTGTCAGTT AATACTGTCGTAGGGAGCAAAG CACTGTGATTAAGATAAGATAAA GATGCTTTAATCTGTACGATTGA AGCATCTTTTTTAAATTTAATATT AATTAATCAGATTTTCAGTTTAA AGTAGGAAAATTTGAAACAGAA AACTTAATAAGTAACACGTAACT TTTATGTGAATTTAGAAAAGCAT AAAAAGGGGTGCACCACCGCGG GTGTATCTACTTTTTATGCTTTTT TATTTAAAATTTAATAGATCAGA CTTGGATCCAGCTGCA	IDT G-block containing the M10 *C. lundense* tandem reporter controlled by the *B. subtilis lysC* promoter for cloning into pDG1661.
thiS31	TACGACGAATTCCAACAGTAATG TTGATCCTTTTAAATAAGTCTGA TAAAATGTGAACTAAATAAAATT TGATATAAATTGCAGGGGAACTA ATTATAGTTCTGAAGGATAAAAC	IDT G-block containing the M11 *C. lundense* tandem reporter controlled by the *B. subtilis lysC* promoter for cloning into pDG1661.

	CTGACCCTTAGAACCTGTCAGTT AATACTGTCGTAGGGAGCAAAG CACTGTGATTAAGATAAGATAAA GATGCTTTAATCTGTACGATTGA AGCATCTTTTTTAAATTTAATATT AATTAATCAGATTTTCAGTTTAA AGTAGGAAAATTTGAAACAGAA AACTTAATAAGTAACACGTAACT TTTATGTGAATTTAGAAAAGCAT AAAAAGGCGTGCACCACCGCGG GTGTATCTACTTTTTATGCTTTTT TATTTAAAATTTAATAGATCAGA CTTGGATCCAGCTGCA	
thiS32	TACGACGAATTCCAACAGTAATG TTGATCC	Forward primer for the PCR amplification of the *C. lundense* tandem reporter construct
thiS33	TGCAGCTGGATCCAAGTCTGATC TATT	Reverse primer for the PCR amplification of the *C. lundense* tandem reporter construct

Chapter Five

Utilizing riboswitches to monitor bacterial physiology during high-throughput small molecule screens

Largely adapted from a manuscript being prepared for publication:

Perkins KP, Atilho RM, Moon M, Breaker RR. Utilizing riboswitches to monitor bacterial physiology during high-throughput small molecule screens.

Summary

Antibiotic drug development is losing the arms race against antibiotic resistance bacteria, importantly to multi-drug resistant gram-negative bacteria. In the past decade, phenotypic drug discovery (PDD) approaches have been regaining popularity as empirical results have shown them to be more successful in discovering novel drugs and new targets over target-based drug discovery (TDD) approaches. PDD can be improved by the development of reporter tools such as riboswitch sensors that help in target identification. In Chapter 5, my colleagues and I demonstrate how a ZTP riboswitch-reporter fusion can be used to identify antibacterial compounds that adversely affect folate metabolism from small-molecule libraries. Using this reporter system in a high-throughput screen (HTS), we identified several sulfonamides with unique chemical modifications that trigger ZTP riboswitch-mediated gene expression. These findings demonstrate that riboswitches can be used as rapid and quantitative indicators that bacterial physiology has been perturbed for the discovery of new antibacterial agents.

Introduction

Since the early 1980s, the number of drugs approved by the FDA have steadily declined as difficulties in antibiotic discovery and development have increased (Ventola, 2015; Gould and Bal 2013; Zheng et al., 2013). This comes at a time when there is a critical need for new treatments against multi-drug resistant (MDR) bacteria and difficult to penetrate gram-negative bacteria (Ventola, 2015; Gould and Bal, 2013). The outlook for antibiotics as a solution to bacterial infections has not been this dire since before the discovery and successful clinical studies of sulfonamides (Bickel, 1988).

Historically, antibiotic discovery methods have been divided into phenotypic versus target based screening. Phenotypic screening typically leverages the power of multiple targets by using a model organism but gains little information on the specific target and mechanism of action (MOA) for a successful hit. Target based screening is typically the opposite because it is limited to testing one known protein target *in vitro* with no information on native bacterial defenses such as cell wall permeability, drug-efflux proteins, or drug-metabolizing enzymes. Although target based screening has dominated the past few decades, empirical evidence shows that phenotypic screening has been more effective in discovering first-in-class small-molecule drugs (Swinney and Anthony, 2011; Kotz 2012). Furthermore, additional technologies have been developed to facilitate identifying the targets of the phenotypic-screen hits (Kotz et al., 2012; Payne and Pompliano, 2007; Laggner, 2011; Guiguemde et al., 2010).

A promising tool for PDD is monitoring intracellular activity by using riboswitch reporter systems. Riboswitches are cis-acting, non-coding, structured RNA aptamers that regulate the expression of the adjoining coding region through either a translational or

transcriptional expression platform upon ligand binding (Henkin 2008; Breaker 2012; Serganov and Nudler 2013). There are known riboswitch classes for 30 fundamental ligands, which include coenzymes, nucleotide derivatives, signaling molecules, amino acids, cell wall precursors and ions (McCown et al. 2017) (**Figure 5-1**). Many of these riboswitches are found in phylogenetically diverse bacteria, which includes bacterial pathogens (Blount et. al., 2006; McCown et al., 2017). Potential pathogen inhibitors have previously been discovered using riboswitch regulated *lacZ* β-galactosidase reporters. For example, riboswitch reporters were used to find small-molecules that target riboswitches as their MOA (Sudarsan et al., 2011; Kirchner et al., 2017; Balibar et al., 2018), or more recently, a fluoride riboswitch reporter was used to find compounds that would enhance fluoride toxicity (Nelson et al., 2015). Due to the nature of riboswitch reporters being simple, modular, and self-contained systems, they are easy to manipulate and implement. The use of riboswitches-reporter systems in antibiotic discovery against gram-positive and -negative bacteria has been aided by a recent review on the phylogenetic distribution of the current riboswitch ligands (McCown et al., 2017).

To identify effective pathway inhibitors and their targets, we propose to use riboswitch reporters in PDD to monitor the disruption of critical bacterial metabolites by small-molecule inhibitors. In several bacterial pathogens, a riboswitch regulates the proper expression of genes that are essential for survival or virulence (Blount et al., 2006). Therefore, the proteins involved in the homeostasis of a riboswitch ligand would make effective antibiotic targets and riboswitches provide a sensor to measure this disruption (Kim et al., 2015). For example, the ZTP riboswitch was discovered to respond to the alarmone, AICAR, that signals for folate deficiency (Bochner and Ames,

1982; Kim et al., 2015). In defined minimal media (M9), the purine and folate pathways are actively synthesizing AICAR and 10-CHO-Tetrahydrfolate (10f-THF) for purine production. To synthesize purines, bacteria need 10f-THF for a transformylase reaction with AICAR, therefore any disruption of 10f-THF causes AICAR to accumulate (Bochner and Ames, 1982; Kim et al., 2015). Increasing concentrations of AICAR causes ZTP-riboswitch mediated upregulation of vital folate biosynthesis genes. This response is triggered by the first, and still extensively used, antibiotic class that target folate metabolism in bacteria (**Fig. 5-2A**). This alarmone response to folate stress can be measured by orthogonally providing a ZTP-riboswitch regulated reporter system, such as the β -galactosidase gene *lacZ* (Kim et al., 2015). This method has been shown to sense the inhibition of DHFS and DHFR, both folate biosynthesis genes (Kim et al., 2015). Since the histidine, purine, and folate biosynthesis pathways all converge at AICAR production, disruption of proteins in any of these pathways could be sensed, allowing for the discovery of novel inhibitors and potentially novel targets. This method leverages the power of multiple targets and biological context of phenotypic screenings while providing information on the list of potential targets and MOA of positive leads.

Herein, my colleagues and I report the development of a high-throughput assay that utilizes a ZTP-riboswitch reporter system cloned into *Escherichia coli* to identify compounds that disrupt purine biosynthesis via folate inhibition. We were able to rapidly screen and quantify antifolate activity for over 120,000 small-molecule compounds. This study demonstrates the effectiveness of riboswitches as sensors that bacterial physiology has been perturbed which is effective in discovering inhibitors and further SAR development.

Results

Certain folate inhibitors trigger ZTP riboswitch-mediated gene expression

To discover novel folate inhibitors and effective targets, we developed a phenotypic screen that used the ZTP riboswitch reporter to provide information on the folate pathway target. A number of phenotypic and target-based methods exist. Current phenotypic assays are not very sensitive requiring death or significant growth reduction for positive-hit detection, however many of these hits will be non-specific nuisance compounds that require a secondary assay to eliminate (Payne el al., 2007). Conversely, target-based methods allow for more sensitive and specific detection of positive-hits with the appropriate assay, however many of these hits will fail later because of issues with intrinsic bacterial resistance mechanisms or because the chosen target is ineffective in modulating bacterial growth or virulence. Therefore, to efficiently screen large chemical libraries for novel small molecule antibiotics that inhibit known or new targets, we propose utilizing riboswitch reporters that specifically monitor the homeostasis of their critical metabolites in a phenotypic assay (Kim et al., 2015). Specifically, we utilized a previously described wild-type (WT) *E. coli* strain in which a ZTP-sensing riboswitch from *Pectobacterium carotovorum* controls the expression of its downstream *lacZ* β-galactosidase (Kim et al., 2015).

To test the effectiveness of the ZTP riboswitch to measure folate disruption by different small molecules, Kevin Perkins tested the reporter assay against several different classes of inhibitors that target the folate pathway (**Figure 5-2B**). The drugs tested can be grouped by their targets, which are dihydropteroate synthase (DHPS), dihydrofolate reductase (DHFR), and thymidylate synthase (TS) (Anderson and Wright,

2014). DHPS is an early step in folate synthesis and has been classically targeted by sulfonamide derivatives (**Figure 5-2A**). Sulfanilamide contain an aniline group that mimics dihydropteroates natural substrate para-aminobenzoic acid (PABA). Unlike PABA, sulfanilamide contains a sulfonamide group instead of a carboxylic acid, therefore inhibiting normal folate synthesis. This inhibition causes a decrease in folate synthesis which subsequently inhibits the formylation reaction of AICAR into IMP resulting in AICAR accumulation that is sensed by the ZTP riboswitch in a concentration dependent manner (Kim et al., 2015).

DHFR inhibitors prevent the reduction of DHF to THF and are used to treated cancer and parasitic disease (**Figure 5-2B**). In our riboswitch reporter assay, we were able to sense DHFR inhibition with trimethoprim, methotrexate, and pyrimethamine at micromolar concentrations without significant cell death occurring (**Figure 5-2C**). This level of sensitivity allows for a robust range of testable concentrations and make riboswitch reporter systems effective tools for high-throughput phenotypic drug screening where cost and efficiency inhibit testing several concentrations. However, permetrexed and cycloguanil did not seem to have antifolate activity at even the highest concentration tested (100 uM) (data not shown). Because we also did not see any significant cell death, it is likely that these inhibiters do not cross the *E. coli* membrane or inhibit so weakly that it would require much higher concentrations, rather than a failure of our assay. This is because they were not optimized for use in *E. coli*. Although cycloguanil binding to the *E. coli* DHFR *in vitro* is known (Srinivasan et al., 2015), the results of our assay implies the barrier to drug efficacy is entry into the cell which is an important obstacle in drug design, especially when targeting gram-negative bacteria. This

highlights the importance of *in vivo* testing for effective and efficient antibiotic drug development.

Kevin Perkins then tested 5-fluorouracil and flucytosine which are TS inhibitors (**Figure 5-2B**). Thymidylate synthase plays an important role in the folate recycling pathway which is critical for cells, however, TS inhibitors do not inhibit folate production or ZTP homeostasis. TS is critical for the production of thymine and thus cell growth and division. While it also recycles THF back into DHF in the folate recycling pathway, other recycling enzymes for THF are known. This is consistent with our data, as no reporter signal was detectable with cells treated with TS inhibitors (**Figure 5-2C**). These inhibitors are not being blocked by the cell as we see cell death at high TS inhibitor concentrations, which is indicative by low OD_{600} signals (data not shown). This data shows that our assay is specific for sensing small molecule inhibitors that are effective at targeting enzymes in the folate pathway, leading to ZTP accumulation, in a whole cell context.

Lastly, Kevin Perkins tested raltitrexed (**Figure 5-2B**), which was reported to inhibit the mammalian enzymes DHFR, GARFT and TS. However, *E. coli* have a TS that has been reported as relatively resistant to raltitrexed (Jackson et al., 1993; Fantz et al., 2000). Therefore, we were curious how raltitrexed would behave under our reporter system. If raltitrexed was targeting DHFR enzymes, we would expect to see an increase in our reporter fluorescence, which our results do not show. Instead, we see a decrease in fluorescence signal compared to WT cells (**Figure 5-2C**). This can occur if raltitrexed is inhibiting an AICAR biosynthesis enzyme in *E. coli* which would lower ZTP levels inside the cell. Fluorescence single begins to decrease even before cell death has

occurred, which is not consistent with the results from other TS inhibitors tested, corroborating previous reports. Therefore, we predict that raltitrexed is inhibiting GARFT or another enzyme in the purine pathway before the formation of AICAR.

Although bactrim, a combination of trimethoprim and sulfamethoxazole, and other antifolates are still used, resistance mutations in the DHFR and DHPS genes have been discovered in gram-positive and gram-negative bacteria since the 1980s (Estrada et al., 2016). Most recently approved compounds are derivatives of compounds already on the market. Two novel compounds are currently in development but these target the same DHFR and DHPS proteins as current antifolates. The ZTP riboswitch appropriately respond to antifolates that disrupt folate production and therefore could lead to novel targets. Moreover, our data allows for parsing of other forms of AICAR disruption. Therefore, Kevin Perkins and I subsequently utilized this reporter strain in a high-throughput screen (HTS) to identify antifolate inhibitors that lead to ZTP accumulation from a chemical library of over 120,000 compounds.

Discovery and validation of high activity antifolate small-molecule hits

HTS assays were performed in M9 media in order for the relevant metabolic proteins involved in folate and purine de novo biosynthesis to be active and therefore targetable. Using the above data, we optimized trimethoprim dissolved in DMSO as a positive control for the HTS. The largest fluorescence signal generated for trimethoprim is at 1 uM, which on average was 4.5-fold higher signal to background, before significant death occurs (OD <0.1). Our base line control were strains grown in the same conditions, but with 1% DMSO, which is the equivalent amount of DMSO added with the compounds at

a final concentration of 10 uM. Over 120,000 compounds were screen and reporter gene expression produced by each compound was analyzed as a percent effect where 0% represents the fluorescence of cells treated with DMSO and 100% represents the fluorescence of cells treated with 1 uM TMP (**Figure 5-3A**).

Hit compounds that yielded a percent effect greater than 12% were chosen for further analysis, which consisted of approximately 523 compounds (**Figure 5-3B**). While some of these compounds had percent effect within the range of noise, we decided to cast a wide net knowing a lot of signals could be low due to low cell count from normal growth inhibition or could be positive hits that could be tested at higher concentrations and later improved by structure activity relationships (SAR). Hit compounds were then further analyzed for auto-fluorescence. The assay was run either without fluorescent agent, 4-MUG, or without cells. If a significant percentage of the original signal was generated in either of these assays the compound was removed from further testing which narrowed our hits down to 109 compounds.

While several compounds seemed to be derivatives of sulfanilamide, there were 28 compounds that had novel chemistries. However, many had substructures that resembled Pan-Assay Interference molecules (PAINs), which are known to cause false positives and have been previously reported to be problematic compounds in drug development studies (Baell and Nissink, 2017; Thorne et al., 2010). Thus, we decided to develop another layer of filtering using a mutated riboswitch reporter strain (**Figure 5-3C**). Previous work has shown that transformed strains carrying mutant riboswitch-reporter constructs that altered highly conserved guanine nucleotides within the aptamer failed to exhibit detectable expression upon TMP treatment. Since the mutated riboswitch-reporter construct is a

single point mutation, it allowed us to keep the entire assay the same and probe for PAINs. If the hits are legitimate, no increase in signal should be produced by compounds that are acting via an AICAR dependent MOA. Out of the 28 unique compounds, there were 26 that produced a fluorescence signal increase upon compound addition. These false positives are most likely due to disruption of normal transcription or translation machinery, which is most likely the case for topoisomerase and other DNA/RNA intercalaters that were positive hits.

Of the remaining sulfanilamide derivatives, ones that were easily obtainable were screened with the mutated riboswitch reporter strain. We noticed that certain compounds produced a similar signal increase with the mutated and WT reporter construct. These compounds are most likely activating gene expression of our reporter system by amplifying any leaky readthrough of the *lacZ* gene via translation or transcription machinery. The observation for this hypothesis is most apparent in the topoisomerases and DNA intercalators that were positive hits in the initial screen but later ruled out. Additionally, a small number of compounds that had low fluorescent signal activity near the percent effect cutoff in the original screen did not increase in a dose-depending matter. It is possible that the compound is targeting neighboring pathways, like the histidine pathway that also produces AICAR, but is easily managed by the cell. However, this hypothesis would require further follow-up experiments that are outside the scope of this study.

Although no novel chemical scaffolds were discovered in this compound library, many sulfonamides were identified from the library screen. In bacteria, sulfonamides act as competitive inhibitors of DHPS by acting as a substrate analogue of PABA (**Figure 5-**

4A). Some of these compounds were pharmaceutical drugs currently being used or previously tested sulfonamides that have been previously shown to have antifolate activity. The discovery of these known antifolates in the screening library are a great positive control that this method efficiently discovers small molecules that target folate metabolism in bacteria. However, within the screen we were able to identify a small number of interesting sulfonamide compounds that deviate from ones currently used (**Figure 5-4B**). These compounds indeed trigger riboswitch reporter gene expression in cells to comparable levels of other known folate inhibitors. Once possibility is that the binding pocket of DHPS is more flexible then previous thought and can accommodate these larger functional groups. It is also very likely that many of this compounds can be broken down inside the cells to its functional form. Further analysis is necessary to determine the precise MOA of this unusual sulfanilamide derivatives.

Discussion

Infections caused by multi-drug resistant (MDR) bacteria challenge many clinicians to find safe and effective treatment regimens for often immunocompromised hosts they target. These bacteria are typically Gram-negative pathogens that contain a cell envelope to limit permeability of small compounds and a large number of efflux systems that make them resistant to many available antimicrobial compounds. These features also hinder the ability to identify novel antibiotics using whole-cell assays, as these compounds are often effuxed out before they can inhibit bacterial growth or viability. By using riboswitches as sensors of bacterial physiology, leads that perturb vital bacterial pathways can be identified by looking for changes in riboswitch reporter activity of sub-inhibitory levels of small molecules compounds.

Herein we describe a high-throughput screen that exploits a ZTP-sensing riboswitch to regulate reporter gene expression in *E. coli.* This screen was successfully used to identify known antifolate inhibitors and unique sulfanilamide derivatives by measuring intracellular ZTP concentrations. The ability of our assay to sense weakly inhibitory compounds or very strongly inhibitory compounds at low concentrations makes it a perfect tool not only for screening but structure activity relationships (SAR). Testing the effects of modifications to the parent sulfanilamide structure can help drug developers understand what chemical motifs improve target binding affinity, membrane permeability and overcome resistance to the parent drug. A caveat in this assay is that it is possible that derivative drugs can be modified to the extent that it can target a new enzyme; however, the mode of action, i.e. purine synthesis inhibition, can still be determined based on reporter “activity”, i.e. ZTP accumulation.

Materials and Methods

Chemicals, DNA oligonucleotides, and bacterial strains. Chemical compounds were purchased from ChemBridge, ChemDiv, LabNetwork, and Sigma-Aldrich. DNA oligonucleotides and double stranded DNA sequences (Supplemental Table S1) were purchased from Sigma-Aldrich or Integrated DNA Technologies. *E. coli* BW25113 strain was obtained from the Coli Genetic Stock Center (CGSC) at Yale University.

Design of riboswitch reporter gene constructs. The WT and mutated ZTP riboswitch reporter constructs in pRS414 plasmids transformed into *E. coli* are the same as reported previously (Kim et al., 2015) with some modifications.

Agar diffusion assays. *E. coli* cultures carrying riboswitch reporter gene constructs were grown overnight ~16 h in defined minimal media (M9) liquid medium. Cultures were then spread on M9 agar plates containing X-gal (80 μg mL^{-1}) and carbenicillin (100 μg mL^{-1}). Autoclaved 6 mm diameter paper discs prepared from 0.35 mm thick pure cellulose chromatography paper (Fisher Scientific) were soaked with 10 μL of compound at specific concentrations and transferred to prepared agar plates. The plates were incubated overnight at 37°C prior to analysis.

Liquid-based β-galactosidase assays. Liquid-based β-galactosidase assays were performed as previously described (Nelson et al., 2015) with some modifications. In short, bacterial cell cultures were grown overnight (~16 h) in M9 with appropriate antibiotics at 37°C with shaking. The next day, 8 μL of overnight culture were then added

to each well in Costar black 96-well clear-bottom plates containing 71 μL of fresh M9 and 1 μL of compound, Trimethoprim (TMP) or DMSO control. The plates were sealed with parafilm, covered with wet paper for humidity, and grown for ~18 hours at 37°C with shaking. Following incubation, the absorbance was taken at 595 nm using a Tecan Infinite M200 PRO microplate reader. Subsequently, 80 μL of Z buffer (60 mM Na_2HPO_4, 40 mM NaH_2PO_4, 10 mM KCl, 1 mM $MgSO_4$) and 40 μL of 1 mg mL^{-1} 4-methylumbelliferyl-β-D-galactopyranoside (4-MUG) (dissolved in 50:50 v/v DMSO:deionized H_2O) were added to each well. The mixture was mildly agitated for mixing and allowed to incubate at room temperature for 15 min. The reaction was stopped using 40 μL of 1 M Na_2CO_3. Excitation and emission values were measured at 360 nm and 460 nm, respectively, using a Tecan Infinite M200 PRO microplate reader. Fluorescence units were calculated as fluorescence intensity divided by the multiplication of total cell density (OD_{595}) and the incubation time in minutes.

High-throughput screening. Screening at the Yale Center for Molecular Discovery was performed largely as described above, with some modifications. Defined minimal media with appropriate antibiotic was added via combidrop in 10 μL volumes to each well of 384-well plates. To each well, 20 nl of a 10 mM solution of each compound tested was added via pintool, followed by the addition of 10 μL of a 1:5 dilution of an overnight culture of *E. coli* containing the ZTP reporter. The reporter cells with TMP or DMSO were included on each plate as controls. Assays testing for auto-fluorescence of compounds omitted cells or 4-MUG from the plate. Bacteria were grown in a humidified incubator for ~18 hours at 37°C with shaking. Reporter gene expression was analyzed by

calculating the percent activity of a compounds fluorescence compared to the DMSO control set at 0% and the TMP control set at 100%.

References

Achari A, Somers DO, Champness JN, Bryant PK, Rosemond J, Stammers DK. 1997. Crystal structure of the anti-bacterial sulfonamide drug target dihydropteroate synthase. *Nat Struct Biol.* **4:**490-497.

Baell JB, Nissink JWM. 2017. Seven Year Itch: Pan-Assay Interference Compounds (PAINS) in 2017 Utility and Limitations. *ACS Chem. Biol*. **13:** 36-44.

Balibar CJ, Villafania A, Barbieri CM, Murgolo N, Roemer T, Wang H, Howe JA. 2018. Validation and Development of an Escherichia coli Riboflavin Pathway Phenotypic Screen Hit as a Small-Molecule Ligand of the Flavin Mononucleotide Riboswitch. *Phenotypic Screening* 19-40.

Bickel MH. 1988. The development of sulfonamides (1932-1938) as a focal point in the history of chemotherapy. *Gesnerus* **45:** 67.

Blount KF, Breaker RR. 2006. Riboswitches as antibacterial drug targets. *Nat. Biotechnol.* **24:** 1558.

Bochner BR, Ames BN. 1982. ZTP (5-amino 4-imidazole carboxamide riboside 50 - triphosphate): a proposed alarmone for 10-formyl-tetrahydrofolate deficiency. *Cell* **29:** 929–937.

Breaker RR. 2012. Riboswitches and the RNA world. *Cold Spring Harb. Perspect. Biol*. **4:** 4.

Estrada A, Wright DL, Anderson AC. 2016 Antibacterial Antifolates: From Development through Resistance to the Next Generation. *Cold Spring Harb Perspect Med*. **6:** a028324

Gould IM, Bal AM. 2013. New antibiotic agents in the pipeline and how they can overcome microbial resistance. *Virulence* **4:** 185-191.

Guiguemde WA, Shelat AA, Bouck D, Duffy S, Crowther GJ, Davis PH, Smithson DC, Connelly M, Clark J, Zhu F, Jiménez-Díaz MB. 2010. Chemical genetics of Plasmodium falciparum. *Nature* **465:** 311.

Henkin TM. 2008. Riboswitch RNAs: using RNA to sense cellular metabolism. *Genes Dev*. **22:** 3383-3390.

Kim PB, Nelson JW, Breaker RR. 2015. An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol. Cell* **57:** 317-328.

Kotz J. 2012. Phenotypic screening, take two. *Science-Business eXchange* **5:** 380-380.

Kirchner M, Schorpp K, Hadian K, Schneider S. 2017. An in vivo high-throughput screening for riboswitch ligands using a reverse reporter gene system. *Sci. Rep*. **7:** 7732.

McCown PJ, Corbino KA, Stav S, Sherlock ME, Breaker RR. 2017. Riboswitch diversity and distribution. *RNA* **23:** 995-1011.

Laggner C, Kokel D, Setola V, Tolia A, Lin H, Irwin JJ, Keiser MJ, Cheung CYJ, Minor Jr DL, Roth BL, Peterson RT. 2012. Chemical informatics and target identification in a zebrafish phenotypic screen. *Nat. Chem. Biol*. **8:** 144-146.

Nelson JW, Plummer MS, Blount KF, Ames TD, Breaker RR. 2015. Small molecule fluoride toxicity agonists. *Chem. Biol*. **22:** 527-534.

Payne DJ, Gwynn MN, Holmes DJ, Pompliano DL. 2007. Drugs for bad bugs: confronting the challenges of antibacterial discovery. *Nat. Rev. Drug Discov*. **6:** 29.

Swinney DC, Anthony J. 2011. How were new medicines discovered?. *Nat. Rev. Drug Discov.* **10:** 507.

Serganov A, Nudler E. 2013. A decade of riboswitches. *Cell* **152:** 17-24.

Srinivasan B, Tonddast-Navaei S, Skolnick J. 2015. Ligand binding studies, preliminary structure–activity relationship and detailed mechanistic characterization of 1-phenyl-6, 6-dimethyl-1, 3, 5-triazine-2, 4-diamine derivatives as inhibitors of Escherichia coli dihydrofolate reductase. *Eur. J. Med. Chem*. **103:** 600-614.

Sudarsan N, Cohen-Chalamish S, Nakamura S, Emilsson GM, Breaker RR. 2005 Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem. Biol*. **12:** 1325–1335.

Thorne N, Auld DS, Inglese J. 2010. Apparent activity in high-throughput screening: origins of compound-dependent assay interference. *Curr. Opin. Chem. Biol* **14:** 315-324.

Ventola CL. 2015. The antibiotic resistance crisis Part 1: causes and threats. *P. T.* **40:** 277-83.

Zheng W, Thorne N, McKew JC. 2013. Phenotypic screens as a renewed approach for drug discovery. *Drug Discov. Today* **18:** 1067-1073.

Figures and Tables

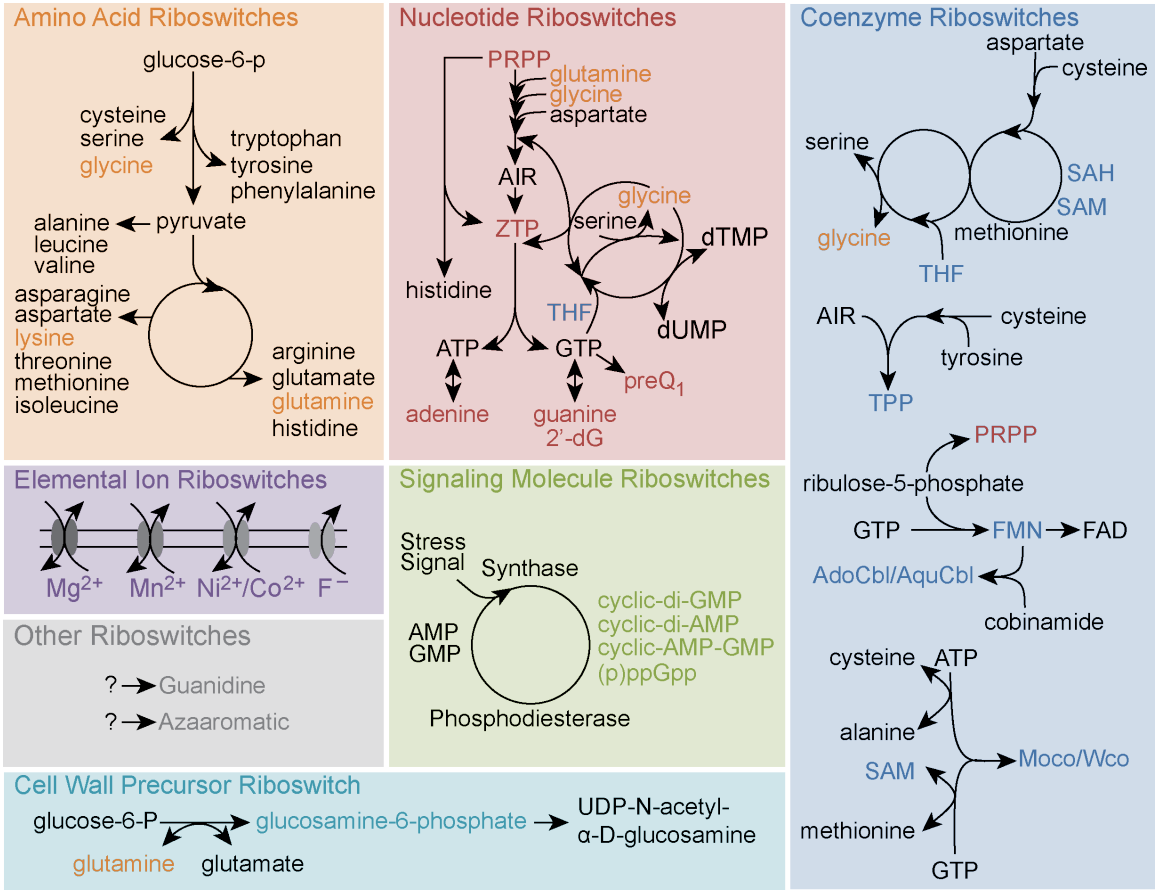

Figure 5-1. Riboswitches control many vital pathways in bacteria. Metabolic and ion-detoxification pathways controlled by riboswitches. Metabolites and ions in color are natural ligands for known riboswitch classes. Arrows indicate expected reaction pathway for metabolite production or ion transportation.

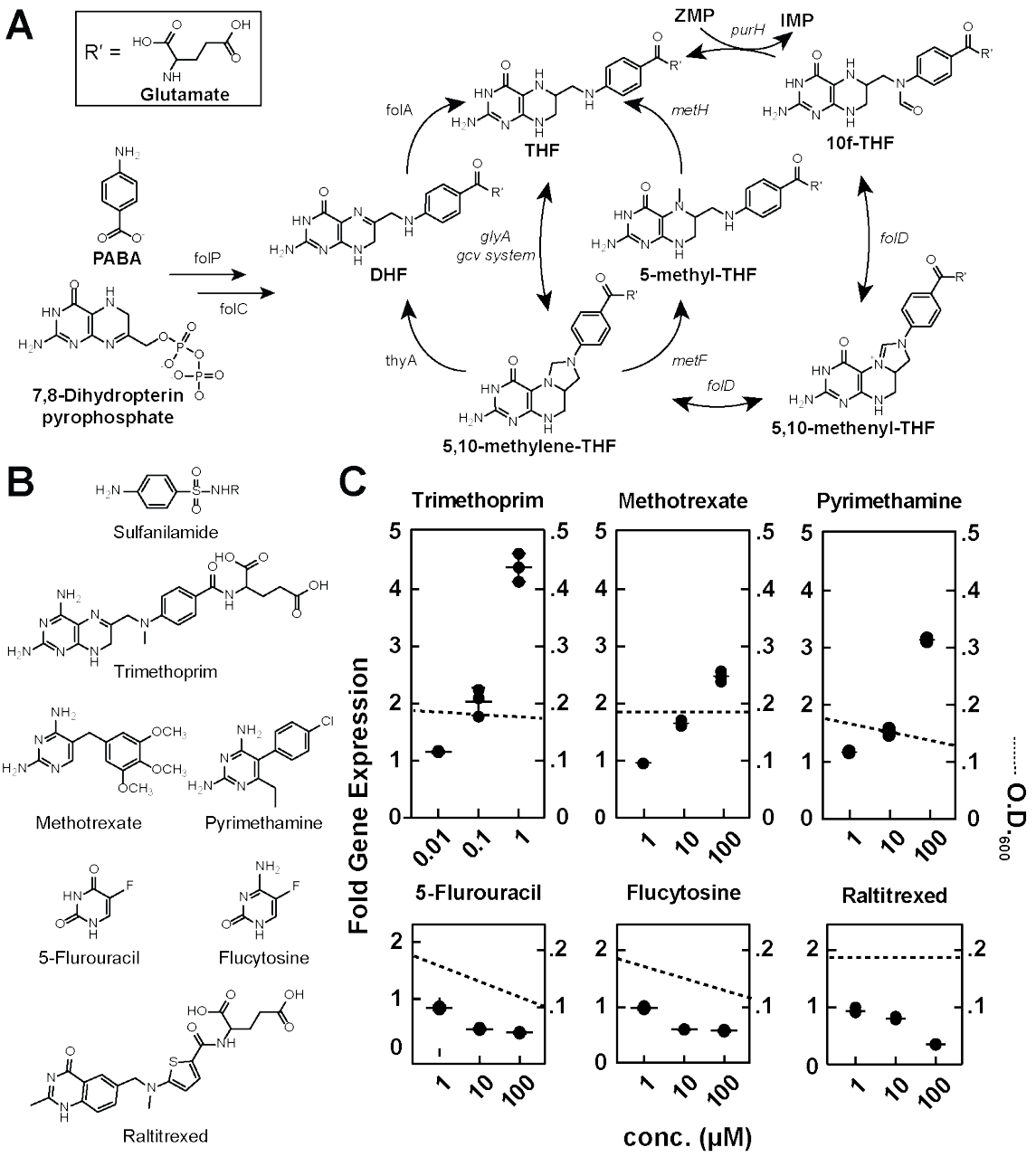

Figure 5-2. Various antifolate drugs activate reporter gene expression. (A) Schematic representation of folate biosynthesis and recycling pathways. Arrows represent expected reaction directionality. (B) Four classes of folate pathway inhibitors, i.e., sulfanilamide derivatives, dihydrofolate reductase inhibitors (trimethoprim, methotrexate, and pyrimethamine), thymidylate synthase inhibitors (5-flurouracil and flucytosine), and raltitrexed. (C) Plot of the relative reporter gene expression levels with increasing

concentrations of known antifolate drugs relative to no treatment in either minimal (M9). Error bars depict SD of data from three independent experiments.

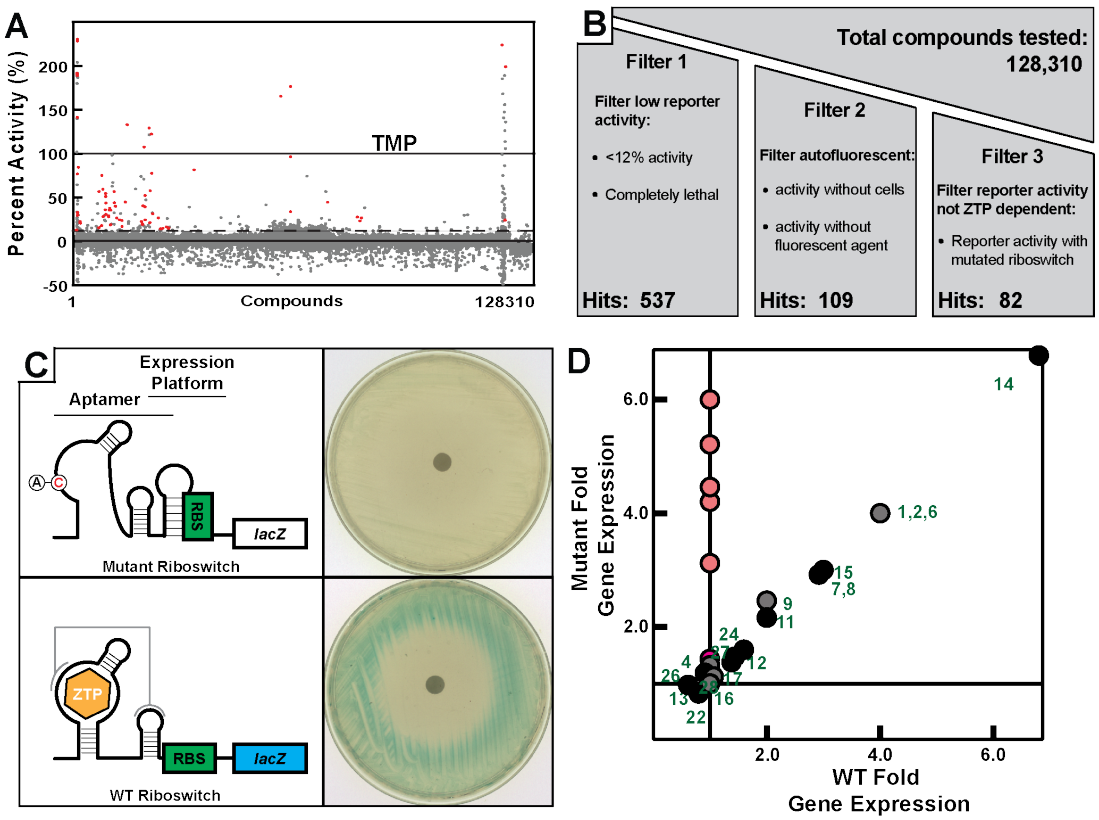

Figure 5-3. High-throughput chemical screen for folate metabolism inhibitors using a ZTP-riboswitch reporter system. (A) A collection of ~128,000 compounds screened at a final concentration of 10 µM for folate inhibition in a *E. coli* ZTP-riboswitch reporter strain. Percent activity is measured by comparing reporter gene expression to cells treated with 1 µM TMP (100 % effect) and cells treated with DMSO (0% effect). Compounds with greater than 12% effect (dashed line) were further characterized. (B) Number of small-molecule hits from **Figure 5-3A** remaining after each stage of filtering. (C) Agar diffusion assays of the *E. coli* Mutant (top) or WT (bottom) ZTP-riboswitch reporter strain. Black discs were spotted with 5 µL of 10 mM TMP. All plates contain M9 agar medium with 100 mg mL^{-1} X-gal. (D) Plot of the relative β-galactosidase activity of WT versus mutant riboswitch reporter strains grown with compounds identified from the HTS at a final concentration of 10 µM.

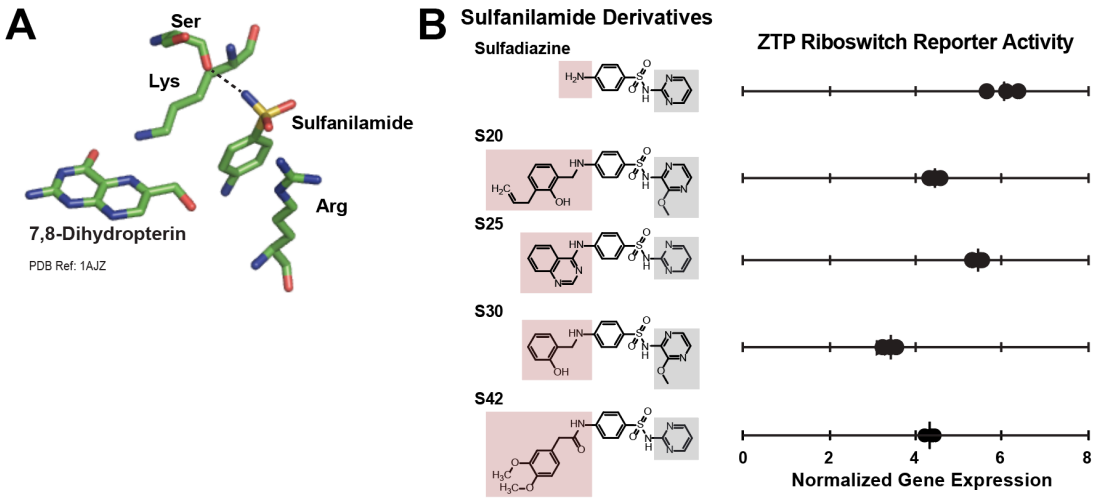

Figure 5-4. Increase in reporter expression for selected hit analogs. (A) Crystal structure of sulfanilamide contacts with 7,8-dihydropterin pyrophosphate and the dyhydropteroate synthase (DHPS) enzyme. Dashed lines identify hydrogen-bonding contacts between serine and the amide of the inhibitor. Hydrophobic interactions occur between the sulfanilamide, lysine, argenine and 7,8-dihydropterin. (Achari et al., 1997) (PDB ID code: 1AJZ). (B) (Right) Structures of selected sulfonilamide derivative hits from the small-molecule screen. Many structures contain an aniline modification that theoretically should disrupt target-inhibition activity. (Left) Individual triplicate results of relative β-galactosidase activity in cells containing the WT ZTP riboswitch reporter construct in the presence of 100 μM of each hit individually compared to no treatment.