

## Abstract

### Monitoring Translation Regulation by Structured RNA Elements

Caroline Marguerite Knotts Focht

2022

Translation initiation across domains of life is a tightly regulated process, and functional RNA motifs are critical regulatory features. This dissertation largely focuses on developing a novel technique with which to quantify the effect these RNA elements have on translation initiation and on unraveling the functional requirements of these structures. I have leveraged RelE's ribosome-dependent endonuclease activity to develop a quantitative assay for translationally regulated mRNAs and validated this method on two bacterial translational riboswitches. I have also demonstrated its applicability to eukaryotic systems by showing that RelE can measure isoform-driven differences in ribosome loading between yeast 5' leader sequences. Moving beyond a gel-based readout, I have integrated RelE cleavage with next generation sequencing to examine the ligand responsiveness of more than 23,000 variants of the *Pseudomonas aeruginosa* Gdm-II riboswitch. Quantitative single and double mutant functional data revealed a finely-tuned expression platform and key positions that temper the switch's sensitivity, dynamic range, and apparent cooperativity. Beyond the comprehensive mutational analysis of known motifs described here, RelE cleavage can also serve as validation of novel regulatory elements identified through high-throughput techniques. Application of this method to RNA thermometers, viral IRESes, variant riboswitches, and T-box RNAs among others will increase our understanding of the biology, evolution, and therapeutic potential of these elements.

Monitoring Translation Regulation by Structured RNA Elements

A Dissertation

Presented to the Faculty of the Graduate School

of

Yale University

In Candidacy for the Degree of

Doctor of Philosophy

by

Caroline Marguerite Knotts Focht

Dissertation Director: Dr. Scott A. Strobel

December 2022

© 2022 by Caroline Focht

All rights reserved.

## Acknowledgements

On my first writing assignment my freshman year of high school, my teacher docked me half a letter grade for not having fully formed thesis statement. That experience has clearly stuck with me to such a degree that I felt compelled to write an entire thesis 13 years later in order to compensate. I hope Ms. Douet is proud of me.

My love for science crystallized in high school when I took biology with Mrs. Messina. She taught me about transcription and translation, and I have a vivid memory of scrawling “I love protein synthesis” at the top of a notes page. To work with ribosomes is truly a dream come true. Mrs. Messina and Mrs. Henderson, my chemistry teacher, challenged and inspired me during school—and they built my confidence in my scientific ability. I’m not sure I would have found my way into biochemistry without them.

In college, courses with Prof. Moeller, Prof. Taylor, and Prof. Wenczewicz helped me realize how much I loved organic chemistry, specifically how that applied to nucleotides and natural products. At the bench, Cheryl Immethun nurtured my burgeoning molecular biology skills, and I could not have asked for a better role model. I would also like to thank my iGEM teammates, mentors, and mentees who taught me not only how to perform assays and experiments, but also how to work with and teach others. I’m immensely grateful to Profs. Pakrasi, Moon, and Jez for allowing me to develop as a scientist in their labs.

When I got to graduate school, I chose to work in the Strobel Lab for its interesting science and, possibly more importantly, its fantastic environment. I’d like to thank Scott for creating and nurturing such a positive environment where people are encouraged to pursue their scientific independence but are never alone along the way. And for starting

an incredible fantasy football tradition—I had never really cared about the NFL before, but I’m genuinely sad my commissioner duties will be coming to an end. Within the Strobel lab, I genuinely can’t thank my mentors—Caroline Reiss (North Caroline), Katie Launer-Felty, Andrew Knappenberger, Chad Torgerson, and Dave Hiller—enough. And of course, I have to shout out Kathryn, Sabrina, Sunitha, Lori, and Alysha for entertaining some of my antics and helping create such a collaborative (and sometimes silly) environment. The Strobel Lab has been such a special place to do research.

The friendships I have made in graduate school will hopefully last a lifetime, and I would like to thank each member of my BQBS cohort for inspiring me every day with their brilliance. I cherish every “Wine and Whine” we had together. And, of course, I have to thank the members of my many intramural and indoor soccer teams who encouraged me to be a better player and teammate every game. I would also like to extend an extra special thank you to my roommate Allison Butt, who definitely had a lot to put up with. Graduate school is filled with highs and lows, and she supported me through all of it with grace and wisdom.

And of course, none of this would have been possible at all without my family. I would not be here today without my parents—and not just in the genetic sense. They have believed in me (and listened to me complain) every step of the way. And to my brothers who gave me middle child syndrome: I’m so happy I got to spend a year in New Haven with Christopher, and I have loved having grad school commiseration phone calls with Michael. Nothing I can say will ever be a sufficient thank you for their endless support, but I hope they know how much I love them (and aren’t finding out from this thesis).

# Table of Contents

<b>1</b>	<b>Introduction.....</b>	<b>1</b>
1.1	Not just a messenger: RNA beyond the central dogma .....	1
1.2	The importance of gene regulation .....	2
1.3	Introduction to translation.....	3
1.3.1	Fundamentals of protein synthesis.....	3
1.3.2	Prokaryotic translation initiation.....	3
1.3.3	Eukaryotic translation initiation.....	4
1.4	RNA Translation Regulation .....	6
1.4.1	Regulation and RNA Structure .....	6
1.4.2	Protein binding sites.....	6
1.4.3	RNA thermometers .....	7
1.4.4	Internal ribosome entry sites.....	9
1.4.5	Riboswitches .....	9
1.5	Kinetic and thermodynamic regulation: a riboswitch case study .....	10
1.6	Methods for monitoring translation initiation.....	14
1.7	RelE: A ribosome-dependent endonuclease .....	15
1.8	Investigating riboregulatory function via Next Generation Sequencing .....	16
1.9	Guanidine riboswitches: windows into unexplored biology.....	17
1.10	Therapeutic potential .....	19
<b>2</b>	<b>Quantifying Translation Initiation with RelE.....</b>	<b>21</b>

2.1	Background.....	21
2.2	Results.....	23
2.2.1	Designing the RelE Cleavage Assay for Translational Riboswitches .....	23
2.2.2	RelE Cleavage Reads Out Ligand-Dependent Riboswitch Ribosome Association.....	25
2.2.3	RelE Cleavage Accurately Demonstrates Switched Riboswitch Specificity.. ..	32
2.2.4	RelE Captures Translational Differences between Eukaryotic 5'-UTR Isoforms.. ..	34
2.3	Discussion.....	38
<b>3</b>	<b>Mutational Analysis of the <i>P. aeruginosa</i> Guanidine-II Riboswitch.....</b>	<b>42</b>
3.1	Background.....	42
3.2	Results.....	46
3.2.1	Developing a Sequencing-Based Assay for Translational Riboswitch Function.. ..	46
3.2.2	Double Mutant Covariation Indicates Structural Features.....	51
3.2.3	High-throughput Mutational Analysis Confirms Features of Bioinformatic Consensus .....	56
3.2.4	Wobble Base Pairs Affect Apparent Sensitivity and Cooperativity .....	62
3.3	Discussion.....	64
<b>4</b>	<b>Conclusions &amp; Future Directions .....</b>	<b>71</b>
<b>5</b>	<b>Materials and Methods.....</b>	<b>75</b>

5.1	Materials and Methods: Quantifying Translation Initiation with RelE	
	Cleavage.....	75
5.1.1	Design of riboswitch constructs.....	75
5.1.2	RNA preparation and labeling .....	75
5.1.3	Translation Initiation and RelE cleavage.....	76
5.1.4	Denaturing PAGE and Analysis .....	77
5.2	Materials and Methods: Guanidine-II Mutational Analysis .....	78
5.2.1	Design of riboswitch constructs.....	78
5.2.2	RNA preparation and labeling .....	78
5.2.3	Translation Initiation and RelE cleavage.....	78
5.2.4	Preparation of RNA for high-throughput sequencing.....	79
5.2.5	High-throughput sequencing.....	80
5.2.6	Analysis of sequencing results.....	81
5.2.7	Calculation of function parameter .....	82
5.2.8	Generating a biochemical consensus .....	82
5.2.9	Modeling Minimum Free Energy Conformations .....	85
5.2.10	DNA oligonucleotides and chemicals.....	85
5.3	Materials and Methods: <i>Can</i> PreQ <sub>1</sub> RelE Cleavage .....	85
5.3.1	RNA Preparation and Labeling.....	85
5.3.2	RelE Cleavage Assay and Analysis .....	85
<b>6</b>	<b>Appendix.....</b>	<b>87</b>
6.1	RNA sequences for gel-based RelE assay .....	87

6.2	DNA sequences used in NGS-based RelE assay .....	88
6.3	NGS Replicate Correlations.....	88
6.4	RNA sequence of <i>Can</i> PreQ <sub>1</sub> riboswitch.....	90
<b>References.....</b>		<b>91</b>

# 1 Introduction

## 1.1 Not just a messenger: RNA beyond the central dogma

The question of inheritance captivated even the earliest ancient philosophers. Families share traits—the “essence” of the parents must therefore be distilled into the offspring. The mechanism of inheritance eluded scientists, however, until the mid-twentieth century. Avery et al. demonstrated that nucleic acids (specifically DNA) constituted organisms’ heritable material [1]. From there, the discovery of the structure of DNA led Francis Crick to propose the “central dogma” of biology: the genetic code stored in DNA is transcribed into messenger RNA (mRNA), which is then translated into a functional protein [2]. This hypothesis relegates RNA to a purely intermediate role, simply a messenger. In that same year, however, Hoagland et al. proposed that transfer RNA (tRNA) functioned as the adaptor responsible for decoding the genetic message into amino acids during protein synthesis [3]. Many additional functional roles of RNA have been uncovered in the decades since, helping to bridge the gap between modern biology and the primordial RNA world.

After the Cech and Altman groups discovered the first examples of catalytic RNAs, Walter Gilbert proposed the RNA world hypothesis. As a macromolecule, RNA has the incredible ability not only to encode and decode genetic information, but also to catalyze chemical reactions ([4]–[6]. Thus, diverse RNA molecules could have constituted the genomes and the molecular machines of the earliest replicating systems. Eventually, organisms expanded their metabolic capabilities through the chemical diversity of amino acids and began storing their genetic information in the more stable

DNA duplex, but relics of those ancient RNAs persist in modern life as ribozymes and riboregulators, such as riboswitches. The ribosome, perhaps the most critical extant ribozyme, diverged as the different domains of life evolved. Controlling ribosome activity as a means of gene regulation, however, unites all three branches of life.

## **1.2 The importance of gene regulation**

The ability to adapt to dynamic circumstances is a critical feature of all lifeforms. Cells must respond to environmental changes such as changing nutrients and temperature. Higher order organisms must alter expression levels during different stages of development as well. Thus, tailoring the composition of the proteome is vital to an organism's survival. Regulation of gene expression occurs at every level, from DNA to synthesized protein. Transcription of DNA into RNA is controlled globally and gene-specifically through a variety of protein and RNA factors. The composition of the transcriptome is also moderated by mRNA stability: factors that increase the half-life of unstable mRNAs also increase the protein output of those transcripts. Translation of mRNA transcripts is similarly regulated by proteins and RNA elements both in *cis* and in *trans*. The sequence and structure of 5' untranslated regions (5' UTRs) in prokaryotes and eukaryotes can effect dramatic changes in translation initiation rates in a transcript-specific way. Once a protein has been synthesized, targeted degradation pathways can control the amount of a given polypeptide in the cell. In this thesis, I will focus on prokaryotic and eukaryotic translation and their regulation by RNA elements.

## **1.3 Introduction to translation**

### **1.3.1 Fundamentals of protein synthesis**

As the domains of life diverged, the protein synthesis players changed, but the game remained the same. Translation consists of 4 phases: initiation, elongation, termination, and recycling. The large and small ribosomal subunits assemble on the mRNA during initiation, and elongation extends the nascent polypeptide chain through subsequent amino acid additions. Termination cleaves the polypeptide from the final tRNA, and recycling disassociates the small and large subunits so that they are available for another round of initiation. These fundamental steps of protein synthesis are conserved throughout domains of life; however, the intricacies— particularly of initiation—differ between prokaryotes and eukaryotes.

### **1.3.2 Prokaryotic translation initiation**

Prokaryotic ribosomes comprise 30S and 50S subunits that assemble into a 70S complex. Initiation also involves three initiation factors (IFs), initiator fMet-tRNA<sup>fMet</sup>, and the mRNA to be translated. The initiator tRNA is distinct from other prokaryotic methionine tRNAs and is charged with formylated methionine (fMet). The IFs are responsible both for assembly of the 70S complex during initiation and for disassociation of the subunits after termination. IF1 and IF3 remain bound to the 30S subunit after dissociation and assist with positioning of the mRNA and initiator tRNA on the 30S subunit (Figure 1.1,A) [7]. Loading of the mRNA onto the ribosome is facilitated by base pairing between the purine-rich Shine-Dalgarno (SD) region in the 5' untranslated region (5'-UTR) of the mRNA and the anti-Shine Dalgarno (ASD) element of the 16S rRNA [8]. The SD element is typically ~7 nucleotides upstream of the start codon and thus helps position the

start codon in the P site. Some mRNAs can correctly initiate without SD elements [9], and some transcripts can even initiate without any 5'-UTR at all, although this is less common [10]. Additional 5'-UTR stand-by sites such as poly-U tracts have been shown to stimulate translation via interactions with ribosomal protein S1 [11]. IF2 directly binds the initiator tRNA, promotes the formation of the 30S Initiation complex (Figure 1.1,A) and facilitates the association of the 50S subunit in a GTP dependent step [7], [12]. This elongation-competent 70S ribosome complex is extremely stable. Initiation is the rate limiting step of bacterial translation and is therefore the most tightly regulated.

### **1.3.3 Eukaryotic translation initiation**

Eukaryotic systems are, in general, more complex than prokaryotic systems, and translation initiation is no exception. The eukaryotic ribosome is larger: the 80S complex is composed of 40S and 60S subunits. Where prokaryotic systems use three IFs, eukaryotic systems require upwards of 13 different factors (eIFs) (Figure 1.1,B) [13]. Even the anatomy of eukaryotic mRNAs is more intricate than their prokaryotic counterparts. Eukaryotic transcripts contain a 7-methylguanosine (7mG) cap on their 5' end and a poly-A tail on their 3' end. eIFs interact with both the 7mG 5' cap and poly-A binding proteins (PABPs) at the 3' end to stimulate translation [14]. Additionally, many genes produce 5'-UTR isoforms of different lengths, and these isoforms contain secondary structures, enhancers, and repressors that affect translation rates. Eukaryotic systems also do not utilize Shine-Dalgarno elements to assist with positioning their start codons. Instead, the sequence context around the start codon (the Kozak sequence) determines the likelihood of initiation at an AUG during scanning [15]. Internal ribosome

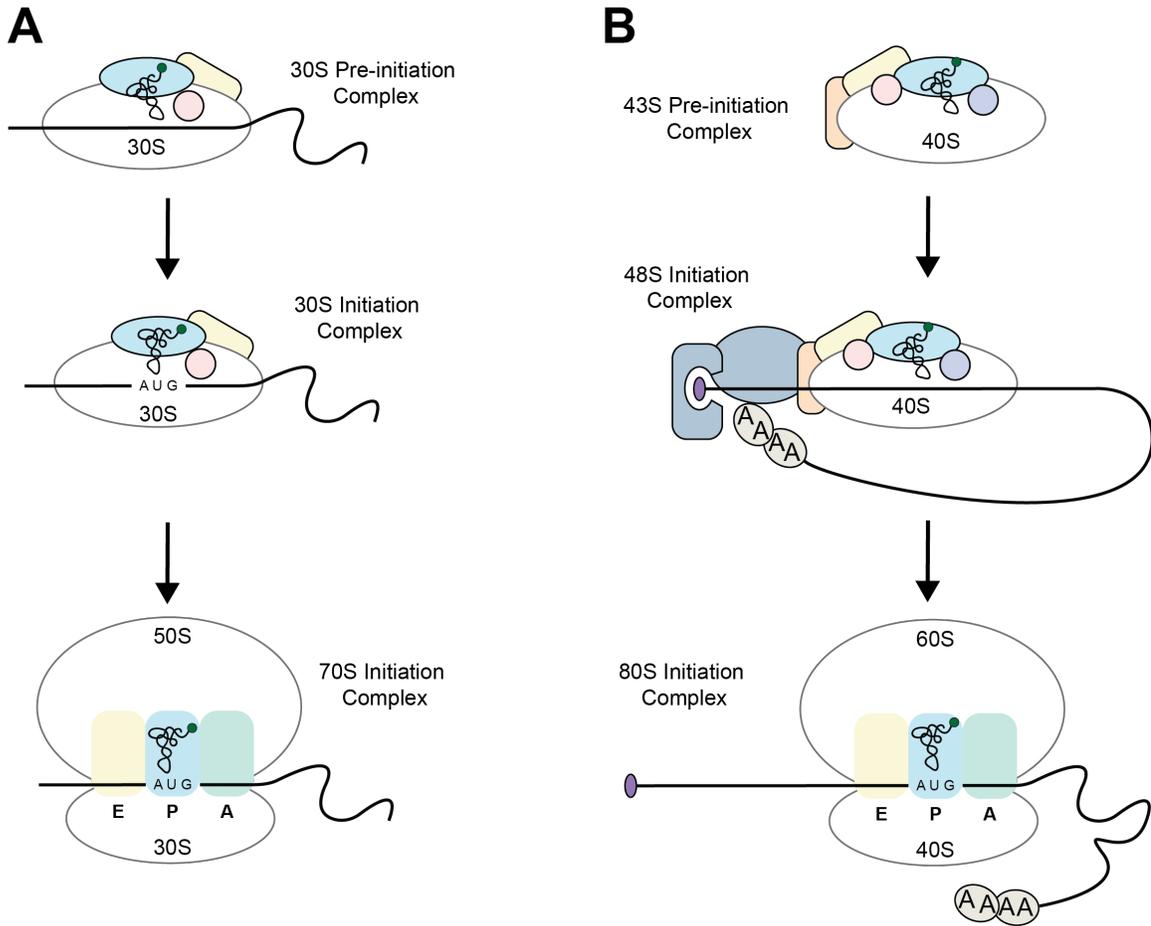


Figure 1.1. Prokaryotic versus eukaryotic translation initiation. A) Prokaryotic translation initiation proceeding from the 30S Pre-initiation complex to the 70S Initiation complex. IF2 shown in light blue; IF3 shown in yellow; IF1 shown in pink. B) Eukaryotic translation initiation proceeding from the 43S Pre-initiation complex to the 80S Initiation complex. eIF1 shown in pink; eIF1a shown in purple; eIF2 shown in blue; eIF3 shown in orange; eIF5 shown in yellow; eIF4F complex shown in grey; PABP shown in beige; 7mG cap shown in purple.

entry sites (IRESes) can be employed by endogenous transcripts as well as viral RNAs to circumvent some steps of eukaryotic translation initiation—particularly its cap dependency [16]. Befitting its intricate nature, initiation is also complexly regulated in eukaryotes both at the cellular and transcript level.

## **1.4 RNA Translation Regulation**

### **1.4.1 Regulation and RNA Structure**

In the initial formulation of the central dogma of biology, RNA was relegated to a purely intermediate role. The discovery that tRNA functioned as the adaptor molecule responsible for decoding the genetic message during protein synthesis, however, revealed the critical importance of RNA structure in essential cellular processes. Beyond its primary sequence, the complex folds of an RNA molecule impart dramatic regulatory capabilities. These functional RNA motifs are particularly important for translation regulation from bacteriophages to eukaryotes. My dissertation largely focuses on developing a novel technique with which to quantify the effect these RNA elements have on translation initiation and on unraveling the functional requirements of these structures. While many modes of RNA-based translation regulation exist, I will focus on the following RNA structural elements.

### **1.4.2 Protein binding sites**

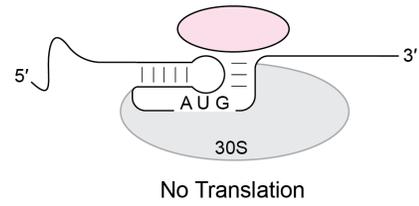
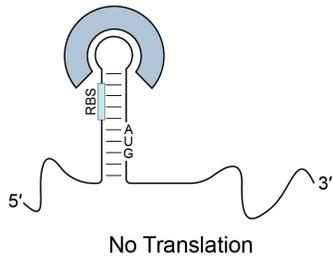
Regulation of translation initiation typically occurs through one of two mechanisms: competition or entrapment. Competition prevents the ribosome from loading onto the mRNA; entrapment prevents a loaded ribosome from accessing an active conformation [17]. In bacteria, protein factors can compete with 30S ribosomes by binding to the ribosome binding site (RBS) or stabilizing RNA structures that occlude the RBS (Figure

1.2). These structures can be simple hairpins or more complex pseudoknots [18]. The MS2 coat protein, for example, stabilizes a small hairpin structure that sequesters the Shine-Dalgarno sequence upstream of the bacteriophage R protein coding sequence [19]. Other protein factors are known to bind proximal to the RBS, and while the 30S subunit may still associate, repressor binding inhibits the 30S subunit from forming a stable initiation complex. The ribosomal S15 protein, for example, regulates its own translation by stabilizing a pseudoknot structure that obscures the start codon [20]. The S15-bound mRNA still makes an SD-aSD interaction with the 16S rRNA and complexes with the 30S subunit, but because the S15-stabilized pseudoknot holds the start codon away from the P site, the 30S subunit is trapped in an unproductive state.

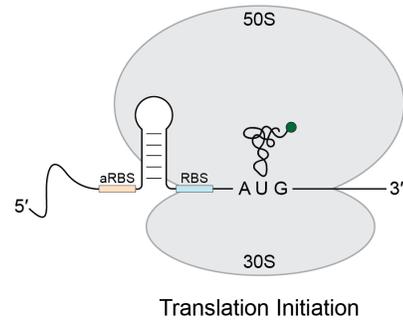
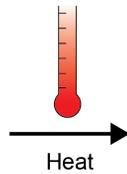
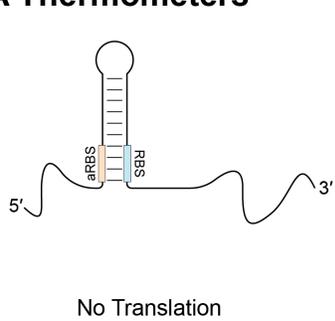
### **1.4.3 RNA thermometers**

RNA thermometers are *cis*-regulatory structural features of prokaryotic 5'-UTRs that regulate translation initiation in response to temperature changes. The thermal stability of these structures is finely calibrated such that changes in temperature result in differences in accessibility to the RBS and thus translation [21]. These elements are frequently used in heat shock response programs and in pathogenic bacteria to control expression of genes only needed during host infection [22]. RNA thermometers typically function as molecular “zippers” in which elevation of temperature melts the helical structure sequestering the RBS. Despite this simple regulatory premise, the structures of RNA thermometers can be extremely complex, comprising hundreds of nucleotides and multiple helices. The precise mechanism of many of these elements is still unknown, and the lack of sequence conservation hinders bioinformatic prediction.

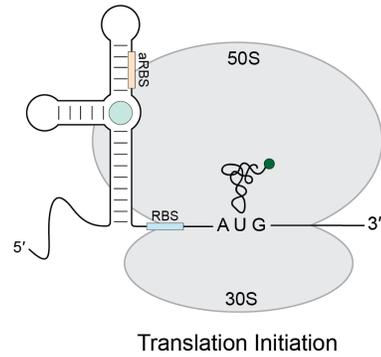
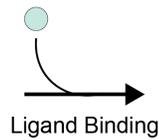
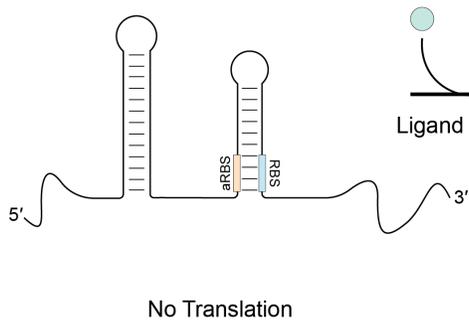
## Protein Binding Sites



## RNA Thermometers



## Riboswitches



## IRESeS

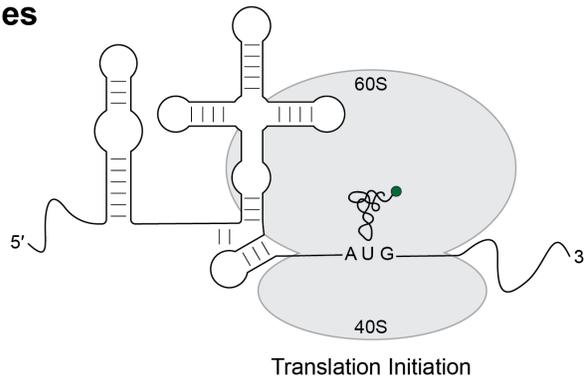


Figure 1.2 RNA Regulators of Translation. Competitive and entrapping RNA binding proteins; RNA thermometer switching with increasing temperature; Riboswitch switching ON in the presence of ligand; Viral IRES loaded onto the eukaryotic 80S ribosome complex

#### **1.4.4 Internal ribosome entry sites**

Endogenous eukaryotic transcripts and viral genomes utilize internal ribosome entry sites. Complex structures in the 5'-UTRs and intergenic regions (IGRs) facilitate translation initiation in the absence of 5' cap structures and without the full complement of eIFs [16]. While structurally diverse, IRESes have been divided into four mechanistic classes [23]. Class 1 and Class 2 IRESes are large structural elements in the 5'-UTR that require almost all eIFs (excluding the cap recognition machinery) to effect initiation. Class 1 and Class 2 are not mechanistically identical, but they generally induce translation of proteins that interfere with cap-dependent translation, thus enhancing their cap-independent initiation. Class 3 IRESes are similarly found exclusively in the 5'-UTRs of viral genes. Most representatives share common structures of multiple stem loops and a pseudoknot, but there is significant sequence and structural variation within the class that may play important regulatory roles [16]. Class 4 IRESes are typically smaller structures found in viral IGRs. These compact pseudoknotted structures do not require many—if any—eIFs and are capable of directly binding to the 40S ribosome. The recently discovered Halastavi árva virus IRES proceeds through an even simpler novel mechanism: direct binding of 80S ribosomes such that binding of a pseudoknot in the P site positions the first codon in the A site for translation [24]. Thus, these diverse structured RNAs effect viral gene expression to various degrees via various mechanisms.

#### **1.4.5 Riboswitches**

Riboswitches are also *cis*-regulatory structural features found predominantly in prokaryotic 5'-UTRs that regulate translation initiation in response to binding of a small molecule effector. To date, more than 55 classes of riboswitches have been identified that

recognize diverse ligands [25]. The ligand and the downstream gene are typically related (e.g. an adenine riboswitch controlling expression of an adenosine deaminase).

Riboswitches can adopt complex folds in order to recognize their ligands sensitively and specifically [26]. Many riboswitch classes contain transcriptional and translational regulators, but some classes comprise exclusively translational examples. Ligand binding to the aptamer domain of the riboswitch produces a conformational change in the expression platform that controls RBS accessibility [27]. These elements can be either ON or OFF switches and may be connected in series with other riboswitch classes to form elegant expression logic gates [28]. While the aptamer regions of riboswitches are amenable to structural studies and bioinformatic prediction, the intrinsically dynamic expression platforms are more poorly understood.

## **1.5 Kinetic and thermodynamic regulation: a riboswitch case study**

Eukaryotic transcription and translation are spatially decoupled: transcription and mRNA maturation in the nucleus is separated from translation occurring in the cytoplasm and endoplasmic reticulum. Thus, the RNA regulatory elements in eukaryotic transcripts may operate in a thermodynamic regime as the elapsed time between processing and export allow RNA structures to achieve conformational equilibrium. Prokaryotic systems, on the other hand, frequently couple transcription and translation. Since these organisms typically lack membrane bound organelles, mRNA is both transcribed and translated in the cytoplasm simultaneously [29]. While the nascent transcript is still being synthesized, ribosomes can access the Shine-Dalgarno sequence and initiate translation. This coupling

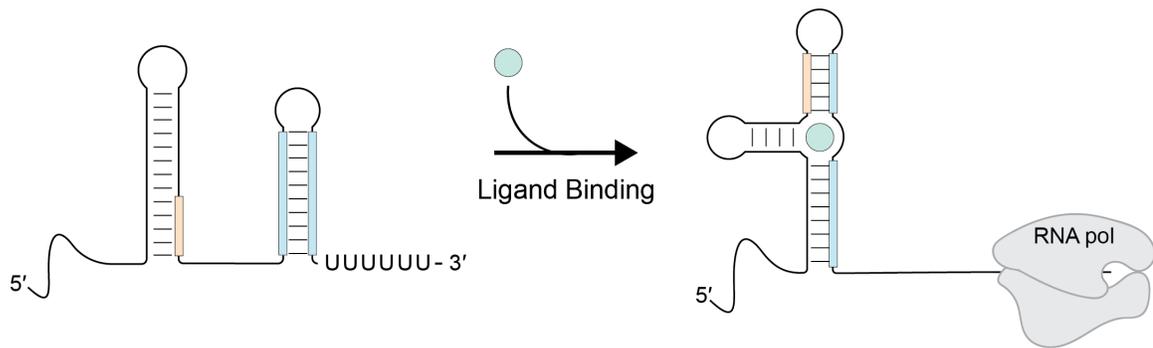


Figure 1.3 Transcriptional ON switch mechanism. Terminator helix shown in light blue, anti-terminator element shown in light orange.

allows regulatory RNA elements in prokaryotic 5'-UTRs to operate within kinetic and thermodynamic regimes.

Kinetically-driven *cis*-RNA regulation is exemplified in transcriptional riboswitch systems, although translational riboswitches are also known to operate kinetically. Like the translational riboswitches discussed in the previous section, transcriptional riboswitches also affect gene expression upon ligand binding. Instead of modulating ribosome association, however, these elements modulate the procession of RNA polymerase through *rho*-dependent or *rho*-independent transcription termination [30]. In the case of *rho*-dependent termination, the conformational change induced by ligand binding exposes *rho* binding sites [31]. Progression of *rho* along the transcript then dissociates the polymerase and effects premature termination [32]. In the case of a *rho*-independent OFF switch, ligand binding stabilizes a terminator hairpin immediately upstream of a polyuridine pause site (Figure 1.3). Formation of the hairpin in conjunction with the weak RNA:DNA duplex within the polymerase results in disassociation of the RNA polymerase from the template and premature termination of the transcript [33]. Since riboswitches are typically short sequences on the order of a few hundred nucleotides, the ligand has a remarkably short window in which to find the aptamer and bind. These elements frequently use internal pause sites and pausing factors such as NusA to moderate the rate of transcription [34]–[37]. In some instances, progression of transcription past a certain point populates binding-incompetent structures such that no ligand binding—and by extension no regulation—is possible. Even with pause sites and NusA as transcriptional speed bumps, the concentration of ligand required to effect a response in transcriptional riboswitch systems is frequently orders of magnitude higher

than the reported affinity of the aptamer. While these elements tailor the kinetic parameters of their system such that switching occurs at the desired concentration and with the desired amplitude, the thermodynamic stability of these different conformational states does still play an important role in their function. Examination of a glycine riboswitch indicated that ligand binding stabilized its alternate conformation by 2-3 kcal/mol [38]; the energetics of the expression platform integrate with the kinetics of ligand binding to produce functional switches.

Translational riboswitches can operate kinetically or thermodynamically. The ThiM riboswitch from *E. coli* inhibits translation initiation upon binding of thiamine pyrophosphate (TPP) [39]. However, progression of transcription too far into the expression platform generates a minimum free energy structure mutually exclusive with the aptamer conformation required for ligand binding [40]. Thus, TPP must bind the riboswitch on a transcriptionally-relevant timescale rather than modulating translation over the entire lifetime of the transcript. The riboswitch systems I study in this thesis operate within a thermodynamic regime. The adenine riboswitch from *Vibrio vulnificus* has previously been identified as a thermodynamic regulator [41]. Binding of the ligand occurs reversibly such that the transcript can respond to changing ligand concentrations over time. In these elements, the energetics of the expression platform helices largely drive the switching decision, and while they certainly affect the dynamic range of the switch (how ON and how OFF the system is in the presence or absence of ligand), these energetics also dictate the concentration at which these elements can respond. However, the available methods to interrogate translational riboswitch function—and question of thermodynamic versus kinetic control—have been largely insufficient.

## 1.6 Methods for monitoring translation initiation

Existing methods of monitoring translation initiation have fallen short when attempting to quantify ligand-dependent changes in translation initiation by riboswitches. For riboswitches predicted to pair the Shine-Dalgarno (SD) with anti Shine-Dalgarno (aSD) elements, helical melting and SD accessibility have been used as proxies for translation initiation in NMR and toehold experiments [42], [43]. Filter binding assays can separate ribosome-associated RNA from free RNA [44]; however, since the assay pulls down any mRNA associated with the ribosome, it is not specific for transcripts that have been properly initiated. Ultracentrifugation has been used to measure the ribosome occupancy of different transcripts, in particular the polysome profile of a given transcript [45], [46], but this cumbersome technique is not readily adaptable to multiple sequences or ligand concentrations. Translation of fluorescent proteins such as green fluorescent protein (GFP) or luciferase can provide a luminescent readout of translation activity [47], [48]. While these measurements can be performed *in vivo* and provide valuable information about riboregulatory behavior in cells, they are frequently subject to high background and noise while also being limited by ligand solubility and membrane transport. Toe printing assays have been used to identify the position of the 30S ribosomal subunit on mRNA transcripts [49], and this technique has provided a binary readout of a ligand-induced difference in 30S subunit association with the *Escherichia coli* thiM riboswitch [50]. Since the method uses reverse transcription as the key readout of ribosome loading, quantification of ligand-dependent changes in translation initiation has not been possible. In 2008, Andreev et al. added an additional dimension to toe printing assays, however, by locating the A site of the initiated ribosome with the bacterial toxin, RelE [51]. In this

thesis, I describe how I have used RelE to develop a facile, quantitative method for monitoring translation initiation.

## **1.7 RelE: A ribosome-dependent endonuclease**

RelE was identified in bacteria as part of a type II toxin-antitoxin system [52]. Under nutritional stress conditions, RelE acts as a global translation inhibitor [53]. RelB, its antitoxin, directly binds and sequesters RelE, thereby preventing its inhibitory activity. RelB is very rapidly degraded by Lon proteases, however; thus, RelE is inhibited only when the cell is capable of synthesizing additional RelB to neutralize its activity. During nutritional stress, RelB is degraded and not adequately replaced, thereby liberating RelE for global translation inhibition.

RelE's inhibitory function derives from its endonucleolytic activity [54].

Interestingly, RelE has no nucleolytic activity free in the cytoplasm—the protein is only active when bound to the ribosome. RelE binds in the A site of initiated 70S ribosomes and cleaves the mRNA between the second and third positions of the codon positioned in the A site, thus preventing further translation. RelE presents moderate codon bias *in vitro* [54], but it cleaves broadly *in vivo* [55], thereby accomplishing indiscriminate translation repression. The preference for a G at the third position of the codon is maintained in *in vivo* studies, however.

RelE homologues have been found across bacterial species and among archaea as well [56], indicating a widespread use for this small toxin. Additionally, RelE has been shown to retain its endonucleolytic activity when bound to the A site of eukaryotic ribosomes as well [51]. This expansion of cleavage activity has been used to locate the A site of initiated ribosomes on eukaryotic and viral transcripts. RelE has also

complemented ribosome profiling studies in bacteria [57], providing further information about the periodicity and coding frame of bacterial transcripts.

## **1.8 Investigating riboregulatory function via Next Generation Sequencing**

The dramatic improvement in DNA sequencing technology over the past decade has revolutionized the possible throughput of nucleic acid-based experiments. Expansions in sequencing depth has allowed for reliable sequencing across genomes, but researchers have also applied this technology to complex synthetic libraries as well. Screening complex pools of sequences for a specific function was pioneered independently by the Gold and Szostak labs in 1990 in their searches for RNA sequences that could selectively bind certain chemical compounds [58], [59]. These systematic evolution of ligands by exponential enrichment (SELEX) experiments separated the functional sequences from a diverse starting pool and enriched them through multiple rounds of selection. Stringent selection and enrichment dramatically reduced the number of sequences contained in the final functional pool such that the sequencing depth was not an issue.

With the advent of greater sequencing power came the expansion of mutational analysis. Instead of whittling down a random complex library of sequences, researchers designed variant libraries of known functional RNAs and performed various levels of mutagenesis [38], [60]–[65]. Instead of discarding non-functional variants through rounds of selection, this approach retained all variants and collected valuable information about the functional landscape of different known motifs. The Greenleaf lab has developed sequencing array methods to probe the effect of mutation on RNA binding protein interactions and ribozyme activity [61], [65]. Using a fluorescent functional readout for

each motif followed by Illumina sequencing of the variant at that position in the array, they can associate functional differences with sequence variants. The Yokobayashi lab leveraged the intrinsic length difference between full length starting material and cleaved products of self-cleaving ribozymes to analyze the effect of mutation on ribozyme function [62], [63]. Previous work in the Strobel lab similarly used the intrinsic length difference between full length and prematurely terminated transcriptional riboswitches to probe how single mutations affected ligand-dependent transcription termination by singlet and tandem glycine riboswitches [38], [64]. While the ribosome does not induce a length difference between translated and untranslated RNAs, in this thesis I explore how RelE cleavage of ribosome-bound RNAs can quantitatively monitor changes in translation initiation among translational riboswitches. By feeding this RelE-induced length difference into a next generation sequencing pipeline as described above, I am able to perform saturating mutagenesis of a translational Guanidine-II riboswitch.

## **1.9 Guanidine riboswitches: windows into unexplored biology**

Riboswitches are often thought of as relics of the RNA world: regulatory systems that control gene expression largely independent of the complex slough of protein factors employed by organisms today. The classes of riboswitches found to date respond to chemically diverse ligands, ranging from ions to vitamin complexes [25]. Riboswitch validation involves pairing the RNA motif with its cognate ligand, and the identity of the downstream gene frequently indicates what that particular motif may bind. In some instances, though, the ligand identity is not obvious due to a scarcity of information about the genes controlled by that motif. For those cases, validation of the riboswitch ligand may in fact reveal important information about the metabolic role of that regulated

protein. Validation of the fluoride riboswitch, for instance, led to the discovery that the downstream gene was in fact a previously unknown fluoride transporter conserved in bacteria [66].

Validation of the *ykkC* riboswitch eluded researchers for over a decade. The downstream genes were largely annotated as urea decarboxylases, but the motif showed no response to urea in in-line probing [67]. The Breaker lab eventually proved that the long-orphan *ykkC* motif selectively responded to the small toxic metabolite guanidine and renamed the motif the Guanidine-I riboswitch class. This finding highlighted that the downstream genes were actually misannotated guanidine decarboxylases, instead.

While guanidinium groups are found within common metabolites (e.g. arginine), guanidine itself was not expected to be a major metabolic player. Since the initial validation of the Guanidine-I riboswitch, however, 3 additional classes of guanidine-responsive riboswitches have been discovered and validated [68]–[71]. The only metabolite with more independent riboswitch classes discovered to date is *s*-adenosylmethionine (SAM), which has six [72]. The prevalence of these guanidine-responsive RNA motifs highlights guanidine's role as a fundamental metabolite. Very little is known about guanidine biology, however. The concentration of guanidine in cells is still unknown, but given its use as a chaotropic agent in molecular biology protocols, cells presumably limit its accumulation beyond the high millimolar range. The  $K_D$  values of the different guanidine riboswitch classes as determined by in-line probing vary, but generally fall in the mid- to high-micromolar range [67]–[71]. These binding constants do not reflect the concentration actually sensed by these riboswitches, however, since the constructs tested frequently omit the expression platform and do not the overall

energetics. The RelE assay I develop in this thesis facilitates the investigation of the functional sensitivity of a translational Guanidine-II riboswitch. Better understanding of the regulatory landscape of these guanidine response elements will enhance our understanding of underappreciated guanidine biology, and the massively parallel translation initiation assay I report here will allow for the functional interrogation of other translational riboregulators.

## **1.10 Therapeutic potential**

The scientific advances of recent years portend a renaissance in RNA therapeutics. The therapeutic potential of RNA has been recognized for decades, but logistical challenges hindered early development [73]. Given its polyanionic nature and inherent instability, RNA presented challenges to packaging and delivery efforts. Since it cannot passively cross the phospholipid bilayer, researchers struggled to get their RNA therapeutics into cells—all while racing the RNA's half-life. For this and other commercial reasons, RNAi-based therapies initially foundered [74]. The development of highly-effective lipid nanoparticles and stabilizing RNA modifications allowed mRNA vaccines to overshadow conventional vaccines in the SARS-CoV-2 pandemic response [75]. While some RNA-based therapies function independently of translation, vaccines in particular require optimization at the protein synthesis level. Mutational analysis of translationally regulatory elements in conjunction with structural studies facilitates engineering and targeting efforts in the development pipeline. Tailoring the protein output of a given RNA is a key step in RNA therapeutic development, as researchers must optimize the efficacy of each dose. Thus, high-throughput methods to identify optimal transcript leader sequences will greatly facilitate the development of these next-generation therapeutics.



## 2 Quantifying Translation Initiation with ReLE

This section is adapted from Focht and Strobel, 2022, an article currently in press in the journal *Nucleic Acids Research*.

### 2.1 Background

Organisms across all domains of life control gene expression at the transcriptional and translational levels. The sequences of the 5' untranslated regions (5'-UTRs) have been shown to alter gene expression in prokaryotic as well as eukaryotic systems [27], [76], [77]. These regulatory mechanisms are prime targets for synthetic biology applications and drug development efforts [78], [79], so the connection between RNA motifs and functional outputs must be better understood.

The well-established transcription termination assay can be used to interrogate how transcription of a bacterial gene is impacted by its 5'-UTR *in vitro* [80]. Using radiolabeled nucleotide triphosphates, the length differences between prematurely truncated and full-length transcripts can be easily resolved via polyacrylamide gel electrophoresis (PAGE). Translational modulators, however, lack such an established assay. The ideal ribosome association assay would take an instantaneous snapshot of ribosome loading. The sensitive, specific assay should be easily scalable to address varied sequences and initiation conditions.

A variety of different approaches have been used to monitor ribosomal association *in vitro* including toeprinting, ultracentrifugation, filter binding, and *in vivo* fluorescent reporters. Toeprinting uses the extension of radiolabeled primers in a reverse transcription reaction to identify the position of a ribosome on a transcript [49]. The bulk

of the ribosome acts as a roadblock, halting the progress of the reverse transcriptase and producing a shorter DNA species. The ribosome is then mapped onto the transcript by comparing the shorter DNA products to the full-length sequence on a sequencing gel. Ultracentrifugation, on the other hand, leverages the size of the ribosome to separate bound from free mRNAs in a sucrose gradient [81]. The laborious and time-intensive method is not practical for small numbers of sequences and varied assay conditions, as each requires a separate sucrose gradient centrifugation. Filter binding assays separate ribosome-bound mRNAs from free mRNAs via ribosome association with a nitrocellulose membrane [44]. Since ribosome association rather than initiation is being measured, however, this approach suffers from non-specific interactions and high background signal. Fluorescent reporters [48] provide valuable *in vivo* data but have limited access to variable condition space.

Here I demonstrate that RelE can be utilized to quantify sequence-driven differences in translation initiation. RelE is a ribosome-dependent endonuclease found in bacteria that is used to globally repress translation under stress conditions [53], [54]. RelE is part of the type II toxin-antitoxin family. It binds and rapidly cleaves mRNAs in the A-site of initiated and elongating ribosomes with a rate constant of  $380 \text{ s}^{-1}$  [82]. In the absence of ribosomes, RelE has no endonucleolytic activity, making it a sensitive and specific measure of mRNA-ribosome association. RelE cleaves the message between the second and third nucleotides and thus defines the mRNA position to nucleotide resolution relative to the ribosomal A-site.

RelE has been previously applied to ribosome profiling studies in bacteria [57]. The precision of RelE cleavage was used to locate the A-site of elongating ribosomes and

refine the position of ribosome footprints. Thus, RelE cleavage was able to reveal important reading frame information among bacterial transcripts. Here I demonstrate that RelE's efficiency and precision can be used to analyze specific mRNA regulatory motifs *in vitro*.

I have leveraged RelE's ribosome-dependent endonuclease activity to develop a quantitative assay for translationally regulated mRNAs. I have validated this method on two translational riboswitches in bacteria with purified components and a commercially available translation system. I have also demonstrated it is applicable to eukaryotic systems by showing it can measure differential ribosome loading based upon yeast 5'-UTR isoforms in a commercially available wheat germ extract. I demonstrate that RelE cleavage is sensitive to subtle sequence changes in both the bacterial and eukaryotic contexts. This RelE cleavage assay can be applied to complex sequence libraries and used to rapidly generate extensive information about the intricacies and functional requirements of regulatory RNAs at the translational level.

## **2.2 Results**

### **2.2.1 Designing the RelE Cleavage Assay for Translational**

#### **Riboswitches**

I developed the following method for assaying translational riboswitch function (Figure 2.1). *In vitro* transcribed and radiolabeled RNA was refolded with varying concentrations of ligand prior to incubation with the ribosomes and initiation factors (IFs). Initiation complexes that form should neither elongate nor release since only the

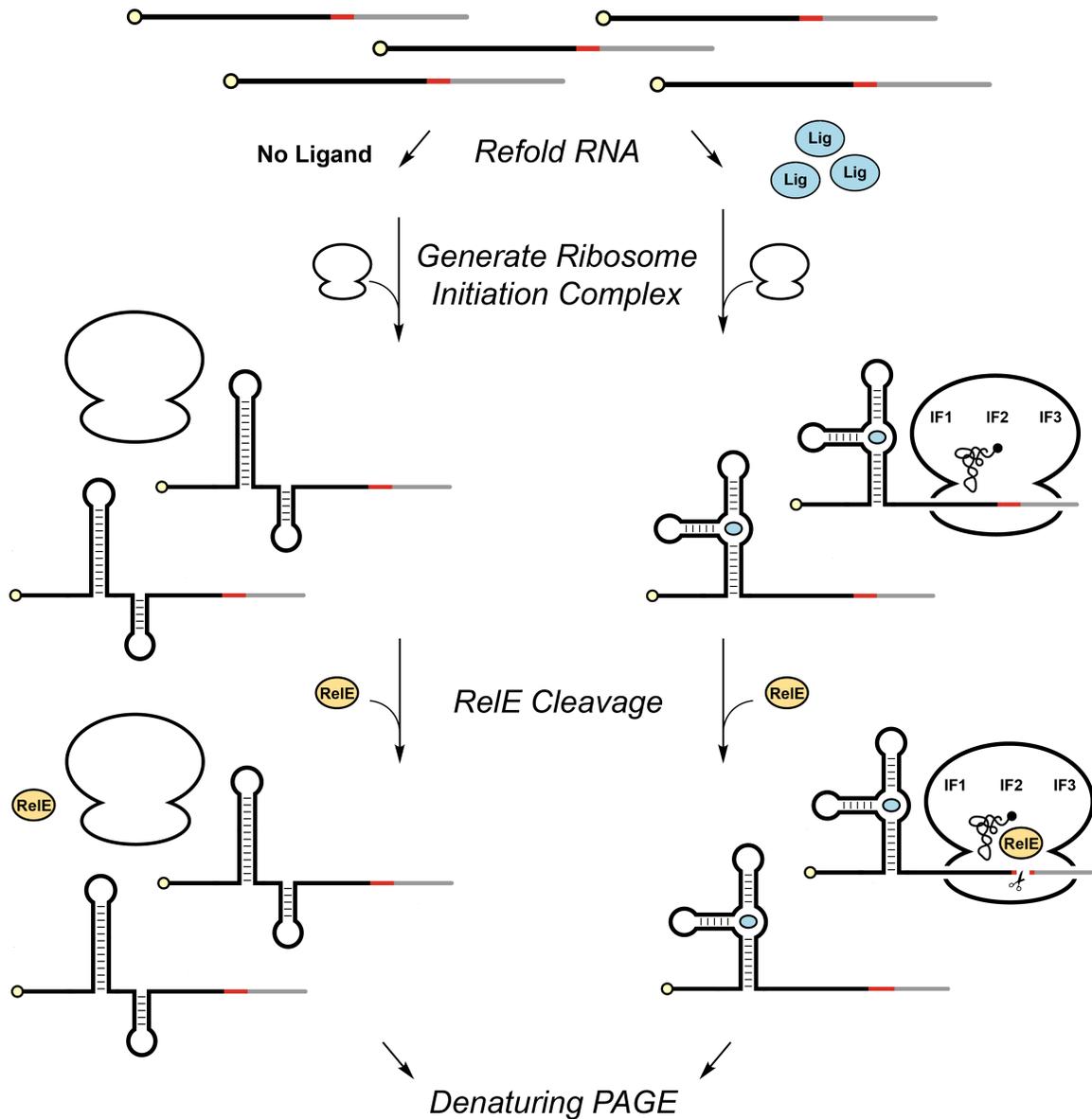


Figure 2.1: Schematic of RelE Cleavage Assay for Translational Riboswitches. Labeled RNA was refolded in the presence or absence of ligand. 5' <sup>32</sup>P label is shown in yellow, site of RelE cleavage in red, and downstream sequence in grey. Refolded RNA was used to form ribosome initiation complexes. RelE was used to cleave ribosome-bound RNAs, and the full-length and cleaved species were separated via denaturing PAGE. The radiolabeled full-length and cleaved RNAs were then quantified across ligand concentrations to determine dose-dependent changes in translation initiation.

fMet-tRNA<sup>fMet</sup> is provided. Following initiation, RelE cleaved the ribosome-bound riboswitch RNA, resulting in a size difference that is ligand-dependent which can be revealed by denaturing PAGE. Percent cleavage of the message was quantified as a function of ligand concentrations to determine dose-dependent changes in translation initiation.

## **2.2.2 RelE Cleavage Reads Out Ligand-Dependent Riboswitch Ribosome**

### **Association**

I first tested the sensitivity of RelE cleavage to ligand-dependent conformational changes in the *Pseudomonas aeruginosa* (*Pae*) Guanidine-II riboswitch. This small motif controls the expression of a multidrug resistance transporter [68] and it is one of over 800 examples of this exclusively translational riboswitch class. The riboswitch features two almost identical hairpins, P1 and P2, and binds two molecules of the positively-charged guanidinium ion through a kissing loop interaction [83], [84] (Figure 2.2A-C). The linker between the two hairpins is poorly conserved. Previous modeling of the *P. aeruginosa* Guanidine-II riboswitch [83] suggests that the 5' and 3' flanking sequences may hybridize to form a P0 helix (Figure 2.2A). This P0 helix is proposed to sequester the Shine-Dalgarno sequence and thus moderates the ribosome's access to the RNA, though there has been no direct biochemical evidence of this mode of regulatory control.

I performed the RelE cleavage assay on the wild-type *Pae* Guanidine-II riboswitch over a range of experimental conditions. Ribosome initiation was IF-dependent (Figure 2.3), and cleavage mapped to the A-site. Cleavage was not detectable in the absence of ribosomes, confirming the ribosome dependence of RelE's endonuclease activity (Figure 2.4A). Similarly, cleavage did not occur in the absence of

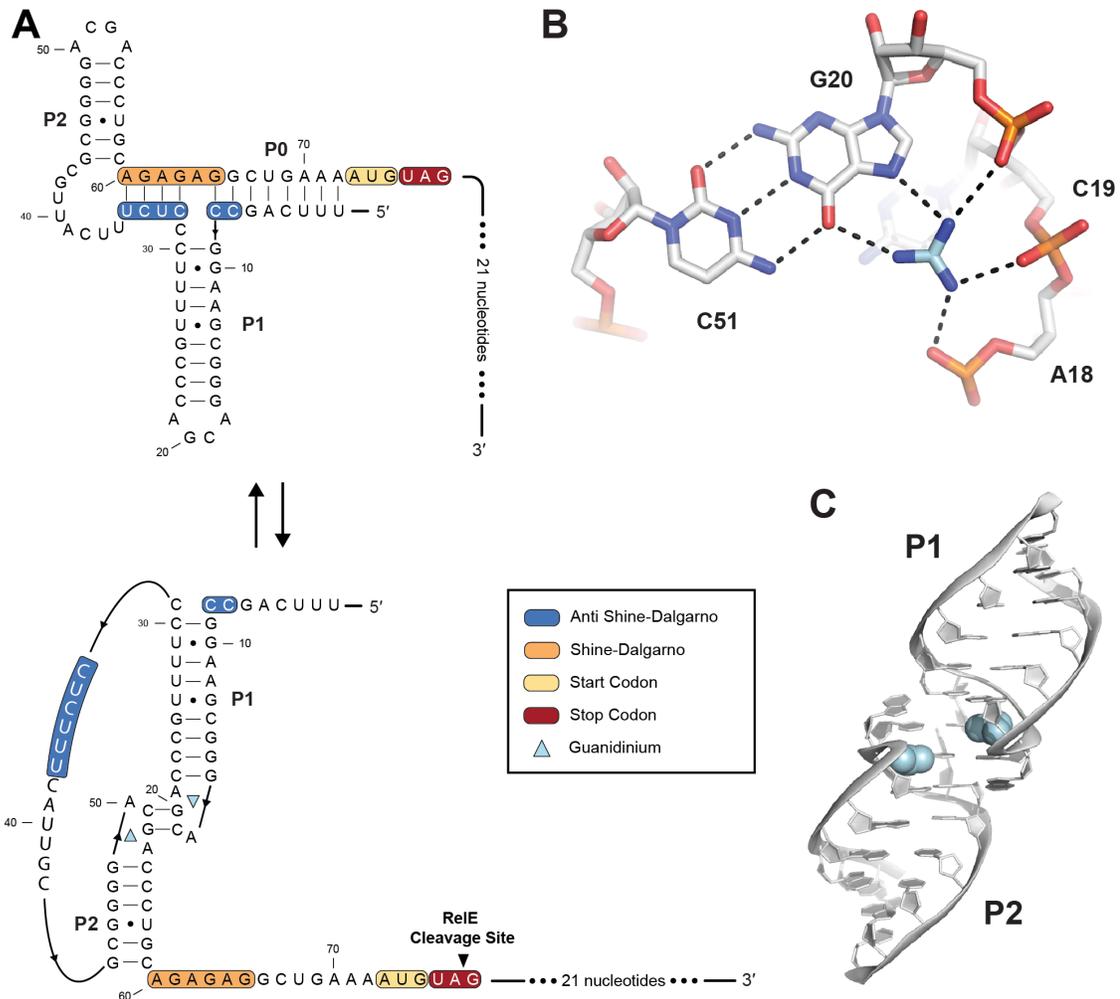


Figure 2.2 ReLE cleavage quantifies ligand-dependent changes in translation initiation A) The secondary structure of the *Pae* Guanidine-II riboswitch in the ON and OFF states. The anti Shine-Dalgarno is shown in blue, the Shine-Dalgarno is shown in orange, the start codon is shown in yellow, and the stop codon for ReLE cleavage is shown in red. B) The P1 binding pocket of the *Pae* Guanidine-II riboswitch. G20 from P1 forms a Watson Crick pair with C51 from P2. The ligand forms hydrogen bonds with the phosphate backbones of A18, C19, and G20 in addition to hydrogen bonds with the Hoogsteen face of G20 C) The crystal structure of the dimerized helices of the *Pae* Guanidine-II riboswitch. The guanidine ligand is shown in light blue (PDB: 5VJ9)

## Time and Factor Dependence of RelE Cleavage

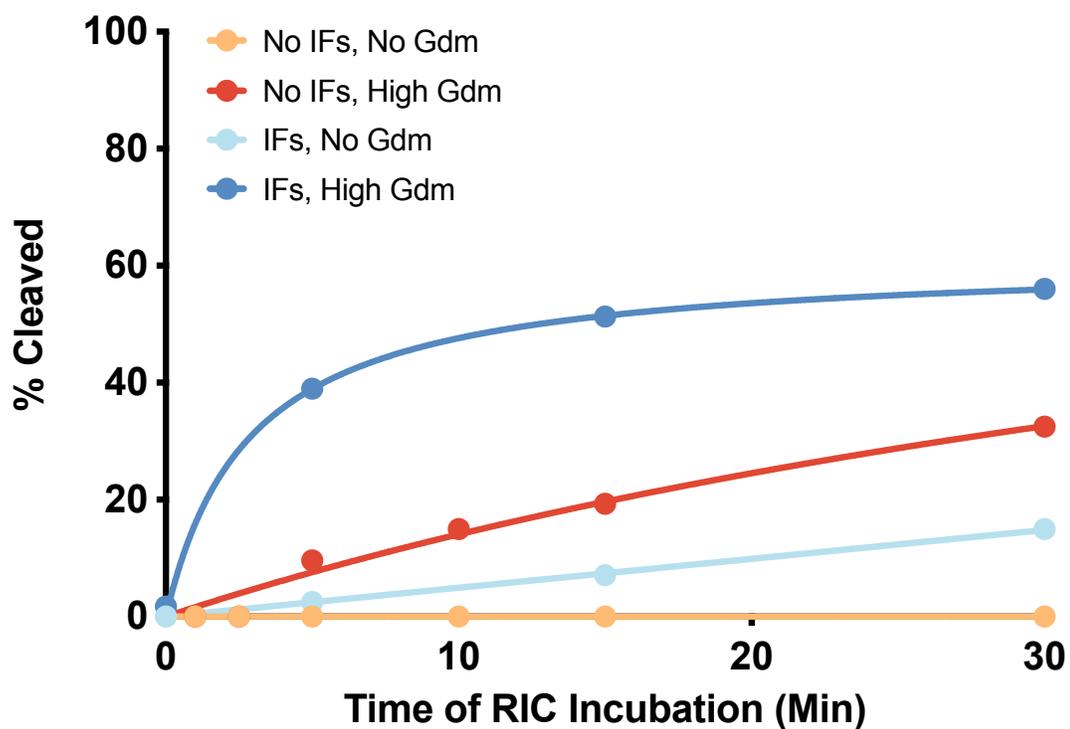


Figure 2.3 Time course of RelE cleavage with and without purified initiation factors at low and high guanidine concentrations. Initiation factors clearly increase rate of ribosome association as read out by RelE cleavage. The data were fit with the following hyperbola:  $Y = Y_{\text{Max}} [X / (K + X)]$

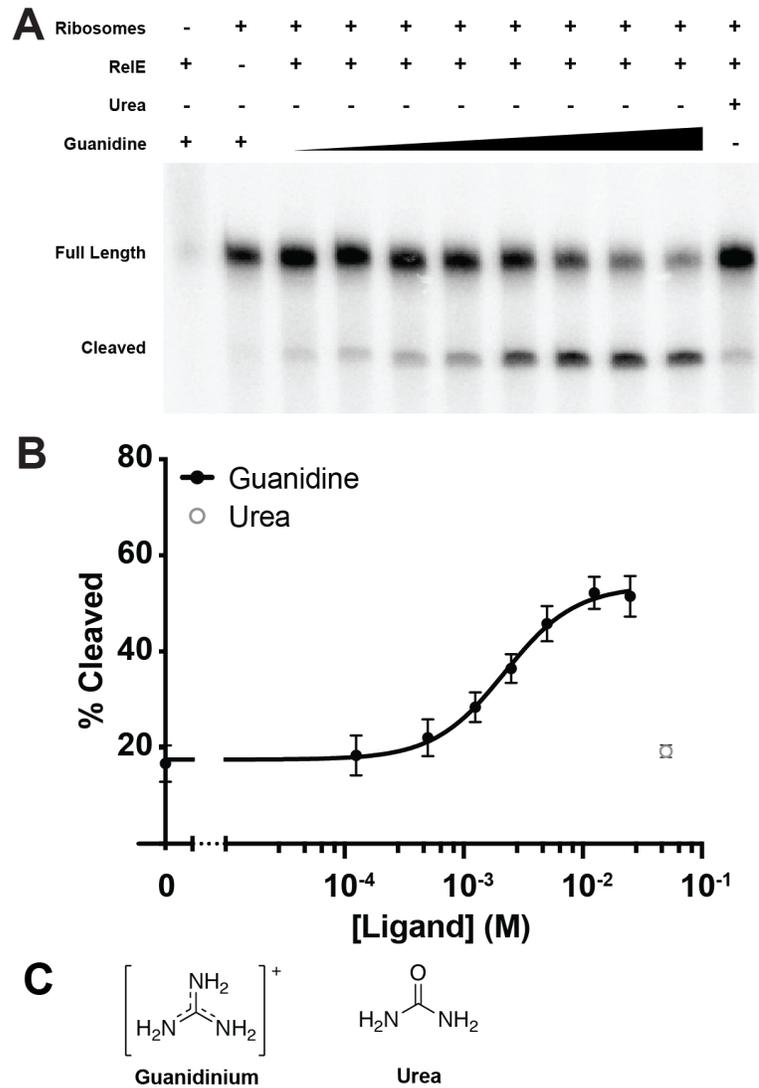


Figure 2.4 RelE cleavage monitors ligand-dependent translation initiation changes in the *P. aeruginosa* Gdm-II riboswitch A) Gel readout of the *Pae* Guanidine-II RelE cleavage. Guanidine concentrations ranged from 0 – 25 mM. B) The ligand response curve of the *Pae* Guanidine-II riboswitch. C) Chemical structures of guanidinium and urea.

RelE. Cleavage was dependent upon the addition of the guanidine ligand and the extent of RelE cleavage was dependent upon the ligand concentration (Figure 2.4A). The percent cleaved at each concentration was fit with Equation 1 since I expected this system to be cooperative [68], which produced measures of the sensitivity ( $K_{1/2}$ ), dynamic range between OFF and ON states (amplitude), and cooperativity (Hill coefficient) of the system (Figure 2.4B). The resulting curve indicates that the *P. aeruginosa* Guanidine-II riboswitch responds to guanidine with an apparent  $K_{1/2}$  of  $2.2 \pm 0.2$  mM, an amplitude of  $36 \pm 4\%$ , and a Hill Coefficient of  $1.5 \pm 0.2$  ( $n = 4$ ). This sensitivity is within the concentration range expected for this metabolite given the reported affinities of other guanidine riboswitch classes and is consistent with the proposed cooperativity of the system [68]–[71]. Cleavage is specific for guanidine binding as no modulation occurs in the presence of urea despite their chemical similarity (Figure 2.4C). A similar guanidine response profile was observed when translation initiation and RelE cleavage were performed on the wild-type *Pae* Guanidine-II riboswitch in a commercially available protein expression system (Figure 2.5).

The RelE assay is also sensitive to sequence-dependent changes in ribosome loading. Mutations were made to the *P. aeruginosa* riboswitch in order to break the riboswitch either constitutively ON or OFF. To break the riboswitch ON, mutations were made to the putative P0 helix to inhibit sequestration of the Shine-Dalgarno sequence (Figure 2.6A). Nucleotides C7, C8, C32, and C34 were mutated to adenines. When subjected to ribosome initiation and RelE cleavage, the ON construct showed no ligand-dependent modulation (Figure 2.6B). Instead, near maximum percent cleaved was maintained across all concentrations. To break the riboswitch OFF,

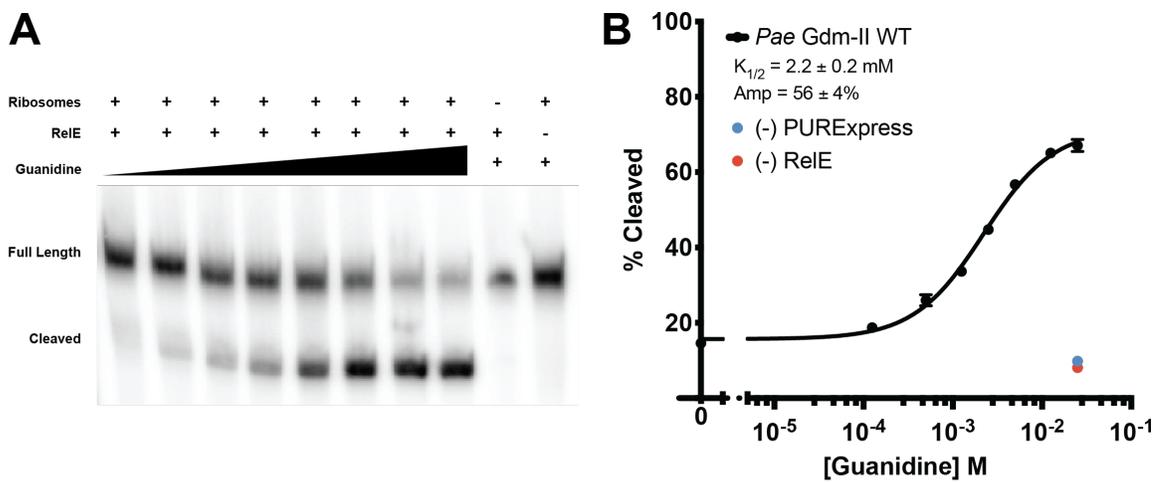


Figure 2.5 RelE cleavage in PURExpress  $\Delta$ RF123. A) Representative gel of RelE cleavage of the *Pae* Gdm-II riboswitch across guanidine concentrations (0 – 25 mM) and with no ribosome and no RelE controls. B) Response profile of the *Pae* Gdm-II riboswitch in PURExpress  $\Delta$ RF123 ( $n = 2$ ). The data were fit with the Hill equation:  $Y = Amp [ X^n / (K^n + X^n)] + Y_{min}$ .

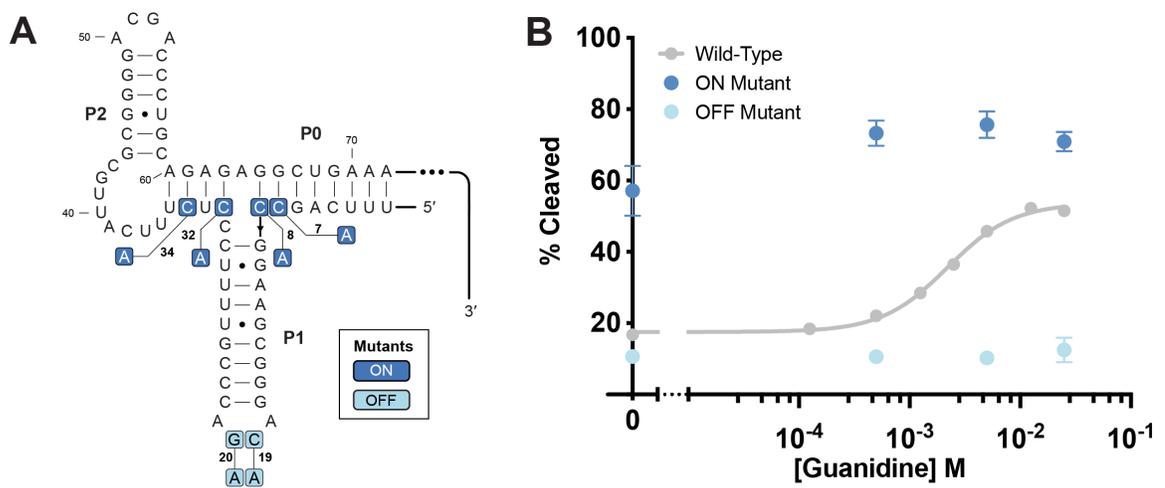


Figure 2.6 RelE cleavage captures mutational functional perturbations. A) Secondary structure indicating the mutations made for ON and OFF constructs. The anti Shine-Dalgarno nucleotides mutated in the ON mutant are shown in dark blue. The binding site nucleotides mutated in the OFF mutant are shown in light blue. B) The ligand responsiveness of the ON and OFF mutants as compared to the WT sequence.

mutations were made to the guanidine binding pocket in L1 to abolish ligand binding (Figure 2.6A). Nucleotides C19 and G20, which directly contact the ligand and zipper the two binding sites together through a kissing loop interaction (Figure 2.3B,C), were mutated to adenines. This mutant similarly lost ligand responsiveness. The minimum percent cleaved was observed across all concentrations (Figure 2.6B). Thus, RelE cleavage sensitively reports functional variations in 5'-UTR sequences. These data suggest that RelE cleavage can be used as a read out for ligand-dependent association of a translational riboswitch with bacterial ribosomes.

### **2.2.3 RelE Cleavage Accurately Demonstrates Switched Riboswitch Specificity**

I further validated the method with the *Vibrio vulnificus* (*Vvu*) adenine-sensing *add* riboswitch. In the presence of adenine, this translational ON switch promotes translation of the downstream adenine deaminase. It is a member of the well-studied purine riboswitch class [85]–[87]. The core aptamer comprises three helices (P1, P2, and P3) with the ligand binding site situated in the multihelix junction formed by J1/2, J2/3, and J3/1 when loops L2 and L3 form a pseudoknot [85] (Figure 2.7A). It was originally identified as guanine-binding switch, but a single C to U mutation in J3/1 switches the specificity to adenine [88].

I performed the RelE cleavage assay on the wild-type *Vvu* adenine riboswitch over a range of adenine and guanine concentrations. The percent cleaved at each concentration was fit with Equation 2 (Figure 2.7A). The wild-type *Vvu* sequence responds to adenine with a  $K_{1/2}$  of  $1.4 \pm 0.7 \mu\text{M}$  and an amplitude of  $28 \pm 2\%$  ( $n = 2$ ). In the presence of guanine, however, the wild-type *Vvu* riboswitch is much less responsive. While the curve cannot be finished due to the solubility limit of guanine,

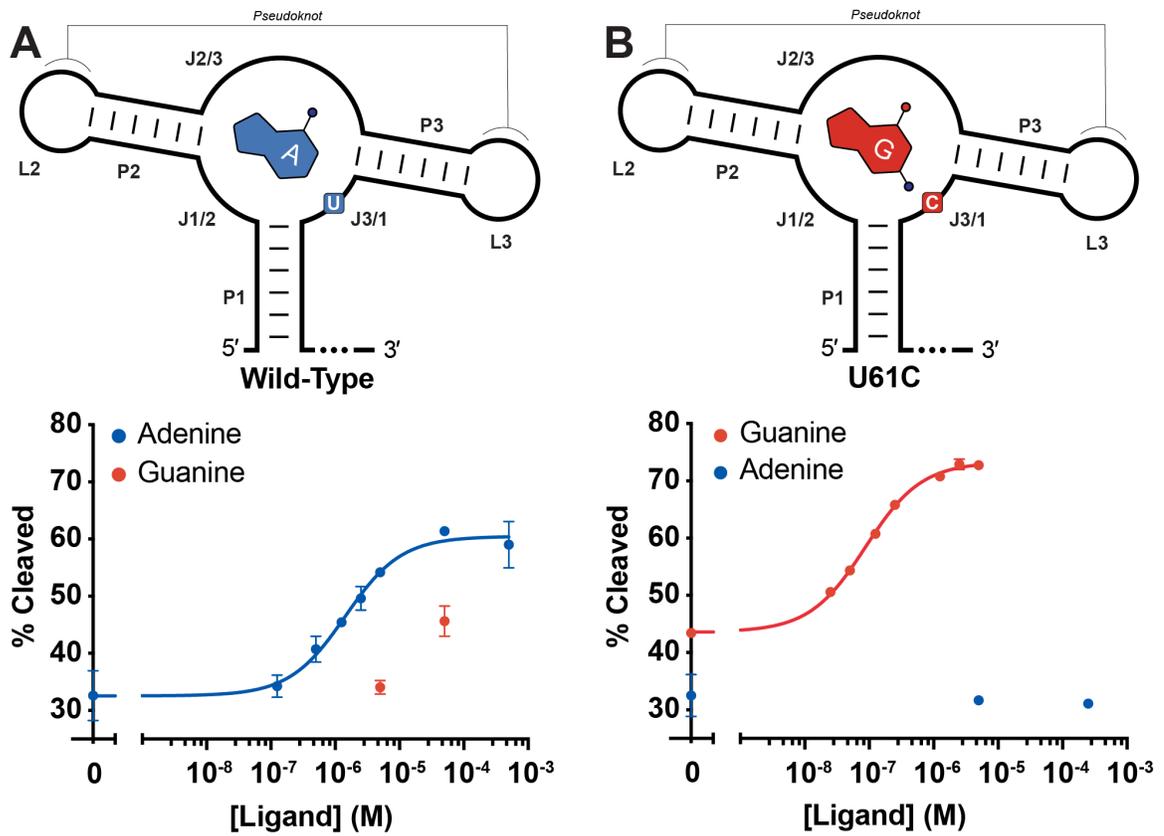


Figure 2.7 ReIE cleavage detects switched riboswitch specificity. A) Secondary structure of the Wild-Type Vvu add adenine riboswitch and the ReIE cleavage data with both adenine and guanine. Adenine and the specificity nucleotide U61 are shown in blue. B) Secondary structure of the U61C Vvu add adenine riboswitch and the ReIE cleavage data with both adenine and guanine. Guanine and the mutated specificity nucleotide U61C are shown in red.

the  $K_{1/2}$  is at least 100-fold weaker (Figure 2.7A). This is consistent with a previous report that the guanine affinity of the ydhL adenine riboswitch is weaker than 10  $\mu\text{M}$  [88]. This promiscuity can be attributed to potential wobble base pairing between the guanine ligand and U61.

I then attempted to switch the ligand specificity of the riboswitch from adenine to guanine. I mutated U61 in our wild-type *Vvu* adenine riboswitch to C in order to allow for canonical pairing between the riboswitch and the guanine ligand. The U61C mutant, as expected, lost all measurable response to adenine (Figure 2.7B). Instead, the U61C *Vvu* riboswitch responded to guanine with a  $K_{1/2}$  of  $87 \pm 3$  nM and an amplitude of  $30 \pm 1\%$  ( $n = 2$ ). RelE thus sensitively reports functional differences in riboswitch specificity.

## **2.2.4 RelE Captures Translational Differences between Eukaryotic 5'-UTR**

### **Isoforms**

I next tested if RelE could be used to monitor the functional differences in translational initiation efficiency between a set of yeast 5'-UTR isoforms. In addition to prokaryotic ribosomes, RelE efficiently cleaves mRNAs loaded onto eukaryotic 80S ribosomes [51]. I therefore adapted my previous workflow to suit eukaryotic translational regulators (Figure 2.8). A recent high-throughput study has reported functional variations in ribosome initiation between thousands of pairs of yeast 5'-UTR isoforms [81]. I selected a pair of isoforms previously identified to be differentially initiated (YGR196C\_40 & \_59) and a pair of isoforms similar in length, but with no reported differences in translation initiation (YML069W\_40 & 57) (Figure 2.9A). After incubating the radiolabeled isoforms in wheat germ extract supplemented with RelE, I quantified

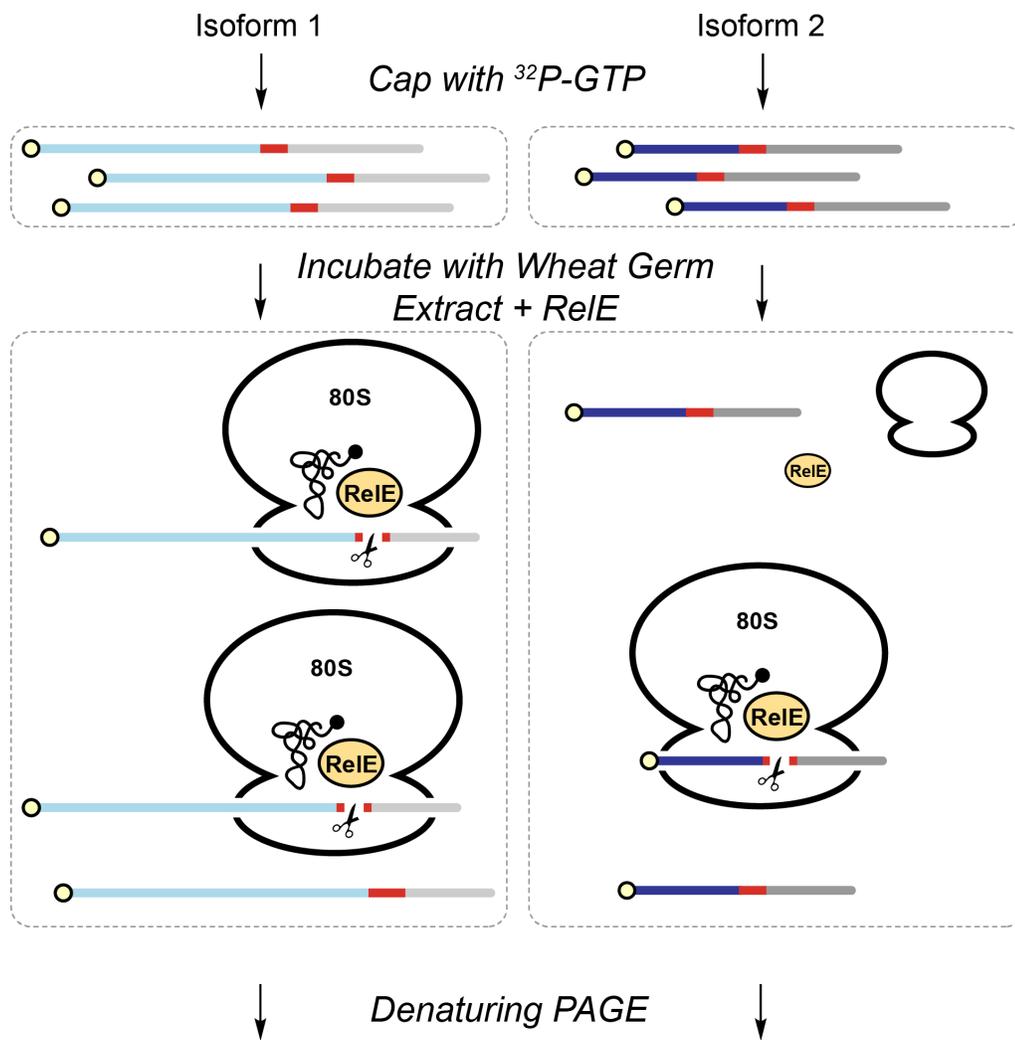


Figure 2.8 ReE captures differential translation initiation in eukaryotic systems A) Schematic of ReE Cleavage Assay for Eukaryotic 5'-UTR Isoforms. 5'  $^{32}\text{P}$  labeled cap structure is shown in yellow. Site of ReE cleavage is shown in red. Downstream sequence is shown in grey.

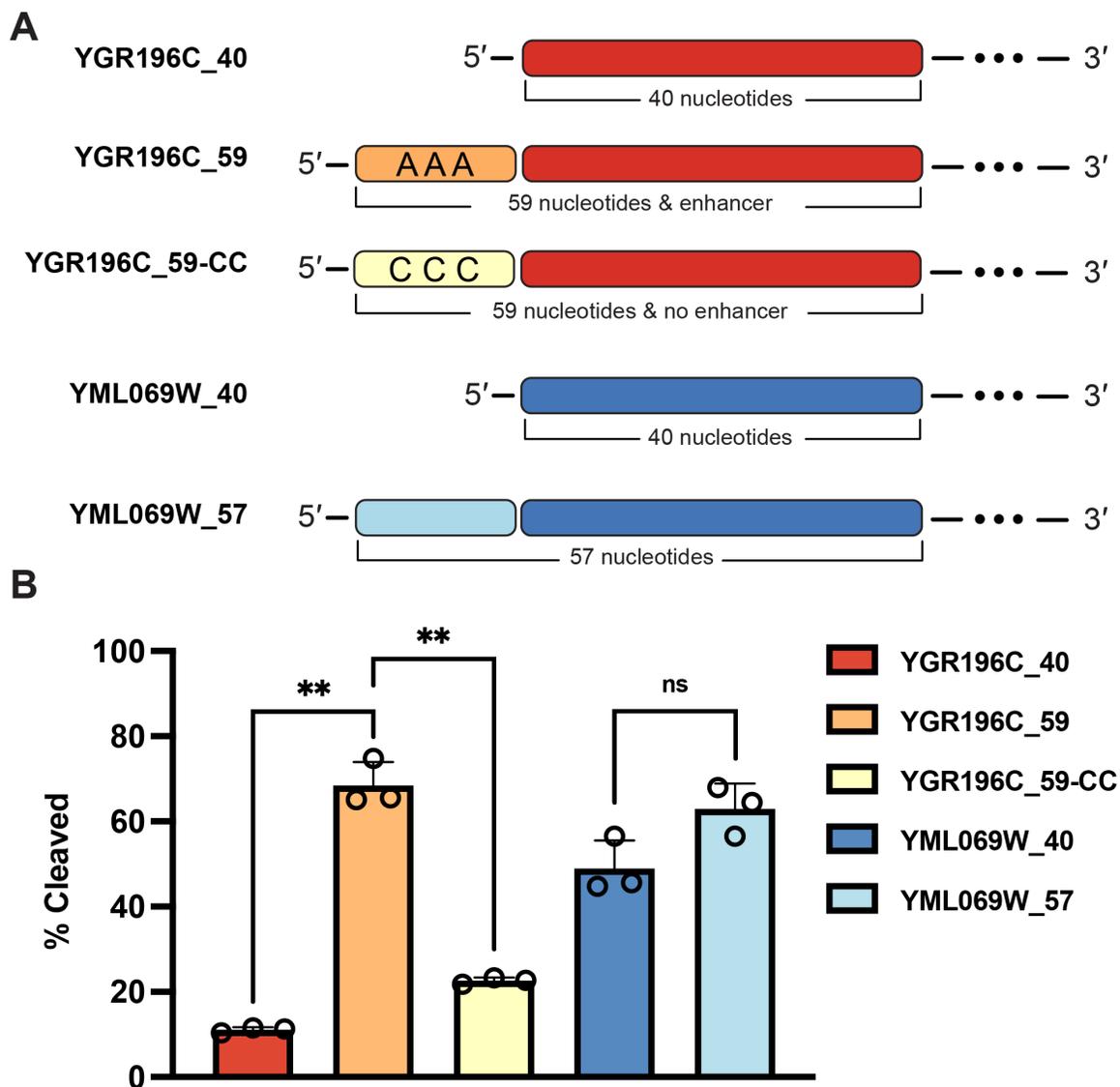


Figure 2.9 A) Representation of the various yeast 5'-UTR isoforms tested. B) RelE cleavage of yeast 5'-UTR isoforms. Two-tailed unpaired t-test with Welch's correlation performed within isoform pairs ( $n = 3$ , \*\*  $p \leq 0.01$ ).

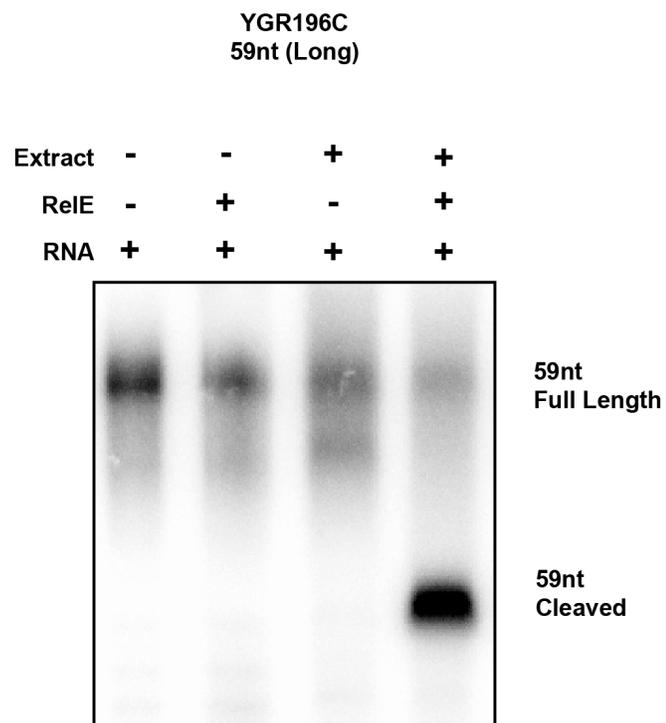


Figure 2.10 Gel separation of full length and cleaved species for the YGR169C\_59 construct alongside wheat germ extract (Promega) and RelE controls.

the percent cleaved for each isoform. RelE cleavage requires both the addition of lysate and the addition of RelE (Figure 2.10). The YGR196C isoform pair shows a significant difference in RelE cleavage ( $p = 0.0024$ ), while the YML069W isoform pair does not ( $p > 0.05$ ) (Figure 2.9B). The long isoform of the YGR196C 5'-UTR (YGR196C\_59) has two previously identified [81] putative enhancer sequences (Figure 2.9A) that are potentially responsible for the observed increase in translation initiation. I mutated these A-rich enhancer sequences to poly-C sequences, and observed a significant decrease in RelE cleavage ( $p = 0.0041$ ) (Figure 2.9B) consistent with reduced translational initiation. Thus, RelE cleavage similarly reports sequence-driven functional differences in translation initiation within a eukaryotic system.

## 2.3 Discussion

I have developed a RelE cleavage assay for the functional interrogation of translational riboswitches and eukaryotic 5'-UTR isoforms. Since sequence and structural motifs have known ties to translation initiation efficiency, this method provides a quick and straightforward way to analyze the effect of various *cis* RNA regulators, their variants, and reaction conditions on ribosomal loading.

The method cleanly resolves ligand-dependent changes in riboswitch-controlled ribosomal loading through the separation of cleaved and full-length species via gel electrophoresis. I demonstrated this for both an adenine riboswitch and a guanidine riboswitch. Structure-guided mutations to the guanidine riboswitch confirmed the assay's sensitivity to sequence-driven changes in translation initiation. RelE cleavage additionally monitors riboswitch specificity as illustrated by switching an adenine riboswitch's ligand preference to guanine.

This assay fulfills the need for a facile, reliable method to monitor the function of translationally controlled riboswitches. Transcriptionally controlled riboswitches have benefited from the well-established transcription termination assay. Not only has this technique allowed for unambiguous confirmation of transcriptional riboswitches' regulatory mechanism, but easy integration with sequencing pipelines has facilitated the study of aptamer-expression platform interaction through high-throughput mutational analysis [38]. Translationally controlled riboswitches, however, have depended on more cumbersome, less direct techniques. Ribosome toeprinting, for example, has been used to validate ribosome association with the *Escherichia coli thiM* thiamine pyrophosphate (TPP) riboswitch [50]. Filter binding has long been used to monitor ribosome-RNA association [89]. Ultracentrifugation has been recently employed as a translation initiation assay for yeast 5'-UTRs [81]. Fluorescent reporters have also been used to validate riboswitch function [48]. RelE provides a snapshot of the ribosome-associated riboswitch population across ligand concentrations while circumventing the need for stringent primer design, laborious washes and centrifugation steps, or fluorescence measurements that are subject to high background levels. This RelE cleavage assay generates quantitative data regarding a riboswitch's sensitivity ( $K_{1/2}$ ) as well as its dynamic range (amplitude) via a standard gel-based readout.

Since the method utilizes *in vitro* transcribed and refolded RNA, the assay is most useful for thermodynamically rather than kinetically driven regulatory mechanisms. While kinetic translational riboswitches have been reported [90], many translationally controlled riboswitches are expected to function thermodynamically so that they can continue their regulatory role over the lifetime of the transcript. The *V. vulnificus add*

adenine riboswitch has been characterized previously as a thermodynamic switch whose function is separate from transcriptional dynamics [41]. I have shown here that the *P. aeruginosa* Guanidine-II riboswitch functions in a similar manner.

The assay also has the potential to refine the resolution of previous work determining ribosome occupancy on various yeast 5' leaders. I have demonstrated that RelE sensitively reports isoform-specific differences in translation initiation, but it also identifies the position of the A site to nucleotide resolution. While my constructs include UAG stop codons as RelE cleavage sites, RelE has been shown to cleave a wide variety of sequences. RelE cleavage could thus be used with endogenous or modestly engineered sequences to determine the contribution of upstream AUGs to the ribosome occupancy of relevant transcripts.

Since the effects of variable conditions can be conveniently examined, this method may also be expanded to other known *cis*-acting 5'-UTR motifs. Variant riboswitches can be screened against various ligands, RNA thermometers can be screened across various temperatures, and enhancer and repressor elements can be screened against various RNA binding proteins.

Beyond condition space, sequence space can also be easily explored. Given its quantitative nature and seamless integration into existing high-throughput pipelines, RelE cleavage can be readily applied to complex libraries of 5' leader sequences. With a sequencing-based readout, this method may be expanded to libraries of endogenous 5' leader isoforms to provide broad insight into translational initiation across the transcriptome. Starting with a mutant library, however, RelE cleavage may also provide fine-grained detail of regulatory motifs such as viral internal ribosome entry sites

(IRESes) through mutational analysis. RelE cleavage can thus reveal a plethora of information about translation initiation in different domains of life.

## 3 Mutational Analysis of the *P. aeruginosa*

### Guanidine-II Riboswitch

I have adapted this chapter from a manuscript currently in progress with Dave Hiller and Scott Strobel. Dave Hiller was instrumental in developing the analytical tools with which we interrogated this massive dataset.

#### 3.1 Background

Riboswitches, RNA motifs found predominantly in the 5' untranslated regions of bacterial mRNAs, control gene expression largely via two mechanisms. Transcriptional riboswitches employ terminator/anti-terminator systems to modulate premature transcription termination. Translational riboswitches control access to the Shine-Dalgarno sequence to temper the mRNA loading onto the ribosome. Extensive structural biology studies have established how these RNA motifs recognize their small molecule effectors [91], [92]; however, the intricacies of how such binding is translated into changes in gene expression remain underexplored.

In the two decades of riboswitch studies to date, our grasp of transcriptional riboswitch function has outstripped that of other riboswitch regulatory mechanisms. Given the facility of transcriptional assays, the functional and folding landscapes of various transcriptional riboswitches have been explored [34], [35], [37], [40], [93]. These studies have revealed crucial interplay between co-transcriptional folding and ligand binding in some systems, indicating kinetic regimes of riboswitch gene regulation (i.e. the  $K_{1/2}$  of the system greatly exceeds the  $K_D$  of the aptamer) in those instances. The detail we have for translational riboswitch regulatory requirements is dramatically sparse

in comparison. We currently model translational riboswitches with OFF states where the Shine-Dalgarno sequence is sequestered and with ON states where the Shine-Dalgarno sequence is free for association with the 30S ribosomal subunit. While some studies have proposed kinetic models for translational riboswitch function [90], other translational riboswitches appear to operate in a thermodynamic regime [94], relying on conformational dynamics over the lifetime of the transcript rather than during the short window of nascent transcription. Recent work has revealed the importance of ribosomal protein S1 in unwinding secondary structure around the ribosome binding site in an adenine riboswitch [43], [95], but more work is needed to study the energetic requirements of these translational regulators. Since some classes of riboswitches comprise only translational representatives, probing the functional landscape of these elements is imperative not only for our fundamental understanding of these regulatory motifs but also for any further engineering efforts.

The Guanidine-II (Gdm-II) riboswitch is an exclusively translational class found predominantly upstream of multidrug resistance transporters [68] and is one of four guanidine-responsive riboswitch families that have been identified in bacteria to date [67]–[71]. Bioinformatically identified as two similar hairpins (P1 and P2) capped by identical ACGR tetraloops and connected by a variable length linker, the Gdm-II riboswitch class binds two guanidinium molecules cooperatively, one in each loop. A Hill coefficient greater than 1 indicates positive cooperativity between multiple binding sites, and in-line probing data for the *Gloeobacter violaceus* Gdm-II riboswitch fit with a Hill coefficient of 1.4 [68]. The bound structure has been solved by X-ray

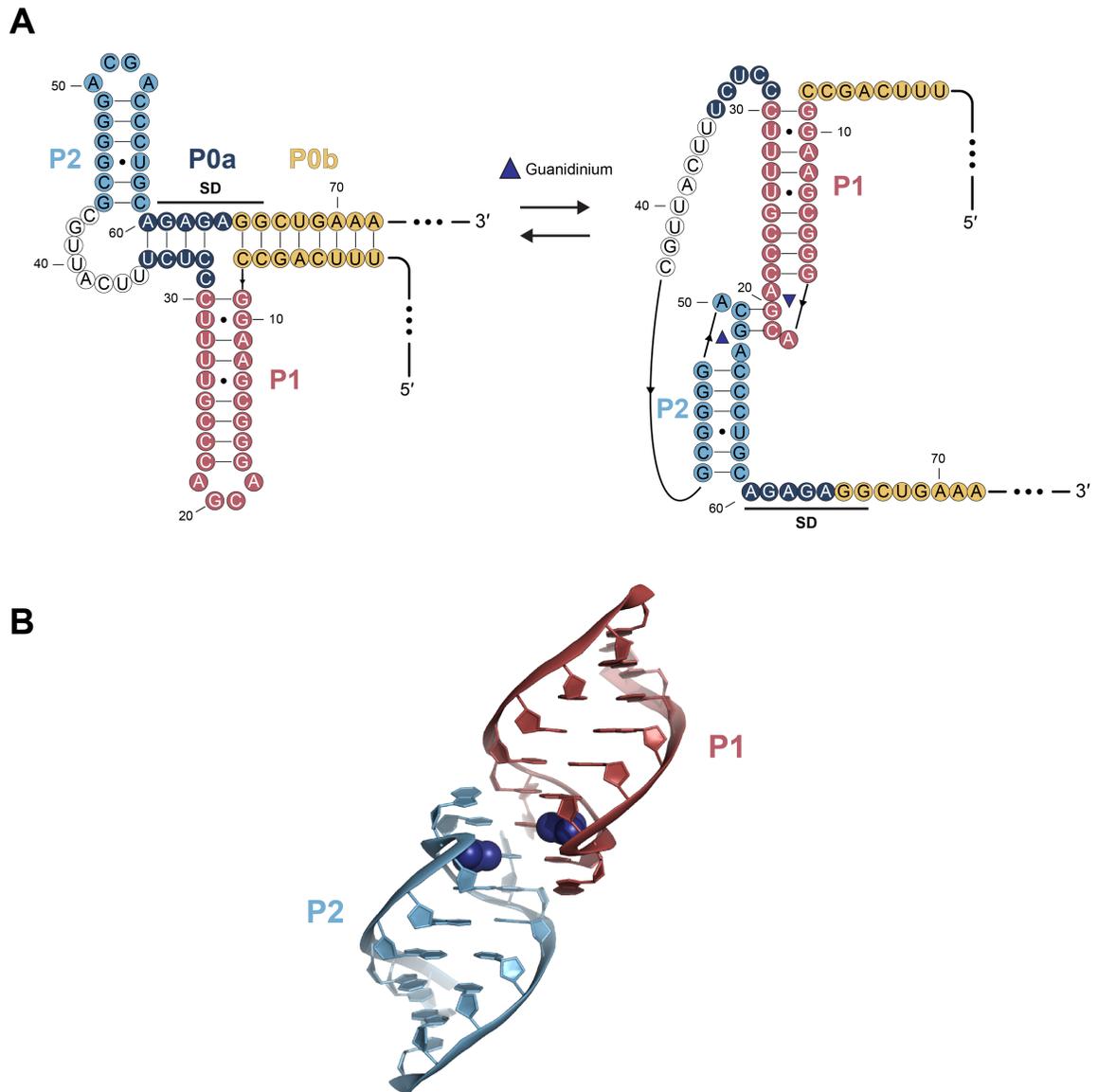


Figure 3.1 A) Secondary structure of the *P. aeruginosa* Gdm-II riboswitch B) Helical dimerization of the *P. aeruginosa* Gdm-II riboswitch

crystallography [83], [84] revealing a kissing loop interaction between the two hairpins. These structures, however, comprise only the hairpin dimers—the linker and flanking regions containing the Shine-Dalgarno sequence (SD) and anti Shine-Dalgarno (aSD) were excluded from structural studies. These dynamic regions are predicted to be integral to the switching behavior of these motifs.

For the *Pseudomonas aeruginosa* Gdm-II riboswitch, the aSD element is held apart from the ribosome binding site when the P1 and P2 helices dimerize in the bound state (Figure 3.1 A,B). No structural information is available for the full-length riboswitch in the OFF state. The OFF state is modeled to include an additional third helix, P0, comprising the hybridization of the 5' tail with the 3' tail (P0b) and the SD with the aSD (P0a) (Figure 3.1A). This P0 helix is proposed to inhibit association of the ribosome in the OFF state, while dimerization of the P1 and P2 hairpins in the ON state pulls the P0 apart so that the ribosome can bind the SD sequence. While the crystal structures have provided crucial information about ligand binding and recognition, analysis of the full sequence of the riboswitch is needed to understand how this RNA element drives ligand-dependent expression modulation.

Recently, many studies have attempted to study regulatory RNA function through high-throughput methods [38], [61], [81], [96]. We have recently reported the use of RelE as an efficient method to monitor ligand-dependent changes in ribosomal initiation. Here we have integrated RelE cleavage with next generation sequencing to examine the ligand responsiveness of more than 23,000 variants of the *Pseudomonas aeruginosa* Gdm-II riboswitch. Quantitative single and double mutant functional data revealed a delicate expression platform and key positions that tune the switch's sensitivity, dynamic

range, and apparent cooperativity. This massively parallel mutational analysis enabled the creation of a biochemical conservation diagram for the *P. aeruginosa* Gdm-II riboswitch, which adds another dimension to the bioinformatic consensus sequence of the Gdm-II riboswitch class. Positions not conserved across the class play pivotal functional roles within the *Pae* sequence context, highlighting the importance of comprehensive functional assays for constructs of interest.

## 3.2 Results

### 3.2.1 Developing a Sequencing-Based Assay for Translational Riboswitch Function

I performed a comprehensive mutational analysis of the *P. aeruginosa* Gdm-II riboswitch using a sequencing-based translation initiation assay. We have previously reported the use of RelE cleavage to quantify changes in ribosome initiation. Since this cleavage produces a measurable length difference between ribosome bound and unbound RNAs, we integrated RelE cleavage with the existing high-throughput pipeline (Figure 3.2) for transcriptional riboswitches, which leverages the intrinsic length difference between full-length and terminated transcripts to report functional changes [38]. I generated a mutant RNA library of the *Pae* Gdm-II riboswitch by *in vitro* transcribing from doped oligos. Positions 1-72 were doped to maximize for double mutations. Translation initiation complexes were formed at various guanidine concentrations, and the RNAs bound to the ribosome were cleaved in the A site by RelE. RNAs were sequenced, and the number of full length and cleaved reads were calculated at each ligand concentration for all 23,220 single and double mutants to generate quantitative

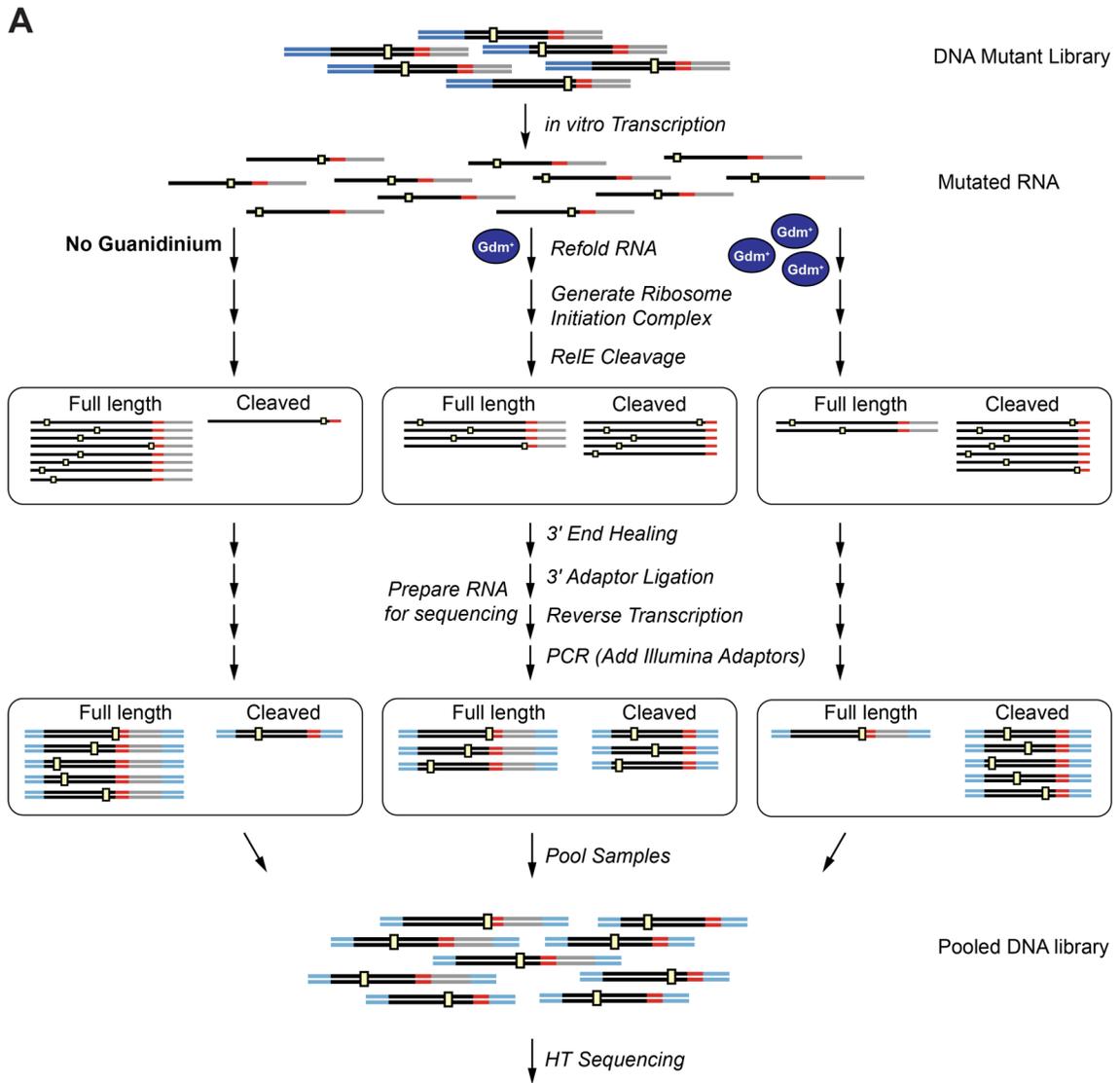


Figure 3.2 Schematic of the high-throughput mutational analysis of RNA translation initiation

ligand-response curves for over 97% of these variants. This made it possible to assess the sensitivity and dynamic range of each sequence.

Since previous analysis indicated this riboswitch is cooperative[68], the percent cleaved at each concentration were fit to the Hill equation (Equation 1). The fit produces an apparent  $K_{1/2}$  that reflects the sensitivity of the riboswitch, an amplitude that reflects the range of the response, and a Hill coefficient that indicates a cooperative response. The trends in the fit data are recapitulated among replicates (Appendix Figures 6.3.1 and 6.3.2).

This dataset reveals that variants within a very limited sequence space can cover almost the entire functional range, with sensitivities spanning 3 orders of magnitude and amplitudes ranging from 0 to 97% (Figure 3.3A,B). Much more variation is seen in the Hill coefficient, which is extremely sensitive to the goodness of the fit, but the distribution is still centered close to the wild-type value (Figure 3.3C). The single and double mutants that could be fit cover a broad range of amplitude and  $K_{1/2}$  values (Figure 3.3E), indicating the tunability of the riboswitch within close sequence space. The  $K_{1/2}$  value maxes out in the mid-millimolar range due to the concentration limits of the assay. Single mutant data indicate that the riboswitch can improve its sensitivity by close to a hundred-fold with a single point mutation (shown in yellow, Figure 3.3E). The least populated space is, as expected, the combination of increased sensitivity and increased amplitude, but we still observe over 3000 variants with  $K_{1/2}$  and amplitude values tighter and greater than wild-type.

This dataset contains almost every possible combination of  $K_{1/2}$  and amplitude within the range of values, but not every combination represents a functional switch. I

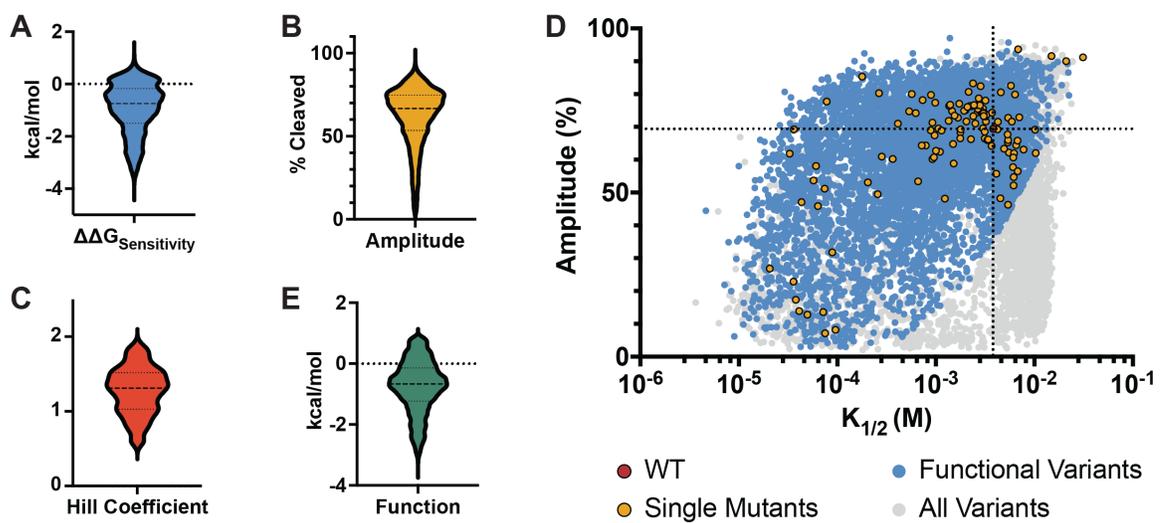


Figure 3.3 A) Distribution of functional  $K_{1/2}$  values ( $n = 7222$ ) B) Distribution of functional amplitude values ( $n = 7222$ ) C) Distribution of single mutant Hill coefficient values ( $n = 127$ ) D) Distribution of function parameter values ( $n = 7222$ ) E) Plot of  $K_{1/2}$  versus amplitude for all variants that were able to be fit (light blue), all functional variants (dark blue), all functional single mutants (yellow), and the wild-type (red).

developed a parameter that combines the apparent  $K_{1/2}$  and amplitude to describe the functionality of each variant (Materials and Methods). Integrating these two values into a single metric of riboswitch function facilitated the analysis of the tens of thousands of variants in this dataset. Riboswitch function depends not only on the concentration of ligand it senses but also on the dynamic range of the expression response. A riboswitch that responds to extremely low concentrations of ligand but barely alters translation initiation is not a functional switch. On the other hand, a riboswitch with a large change in translation initiation that only responds at physiologically irrelevant concentrations of ligand is similarly not functional. Thus, a riboswitch functional parameter requires the integration of the  $K_{1/2}$  and the amplitude. To do this, I transformed the apparent  $K_{1/2}$  into a free energy ( $\Delta G_{\text{sensitivity}}$ ) and the amplitude into a pseudo-energy ( $\Delta G_{\text{amplitude}}$ ). I compared each variant value to the wild-type to create  $\Delta\Delta G$  values. The  $\Delta\Delta G_{\text{sensitivity}}$  and the  $\Delta\Delta G_{\text{amplitude}}$  are similar in magnitude and were weighted equally. The function parameter is thus centered around a wild-type value of 0. Variants that function better than wild-type (tighter  $K_{1/2}$ , larger amplitude) have function parameters less than 0, and variants that function worse than wild-type (weaker  $K_{1/2}$ , smaller amplitude) have function parameters greater than 0.

The distribution of the function parameter is skewed toward values less than zero since the most broken mutants are excluded due to poor fits (Figure 3.3D). Imposing a function cutoff, we separated functional switches (dark blue in Figure 3.3E) from the non-functional (grey in Figure 3.3E). ~50% of the single mutants retain function (yellow in Figure 3.3E). The functional switches maintain the range of amplitude and  $K_{1/2}$  values, while eliminating those variants with both poor amplitudes and weak sensitivities. The

functional space achievable by single mutants is similar to that of the double mutants. Thus, the functional landscape of this sequence is broad and rugged, with steep peaks and deep valleys within reach of a single mutation, and the double mutant data allow us to probe that landscape further by illuminating interactions within the riboswitch.

### **3.2.2 Double Mutant Covariation Indicates Structural Features**

The functionality of each double mutant variant reveals covariation in the P1, P2, and P0b helices. I calculated the difference between the observed and expected function of each double mutant (Materials and Methods), where the expected value is the sum of the function parameters for the relevant single mutations. I plotted these differences as a heat map to identify possible interactions in the riboswitch (Figure 3.4A). Negative values indicate favorable interactions between two mutations in the ON state, and positive values indicate interactions unfavorable to the ON state.

Diagonals of negative values reveal covariation in the P1 and P2 helices (Figure 3.4B&D). Covariation in the P2 helix presents as a single dark blue diagonal, while the P1 helical trend is more complex. We see the ability of slippage in P1 as illustrated by the parallel diagonals (Figure 3.4B). The riboswitch appears to bulge out U26 and maintain pairing of the stem (Figure 3.4B). Replacement of the G-U wobble between G10-U29 with a G-C pair between G10-C30 may energetically compensate for the bulged nucleotide. Evidence of mutation-induced structural rearrangements has previously been seen in mutational analysis of the *glmS* ribozyme [65], but these data suggest that bulging U26 may occur as a natural alternative conformation in the wild-type switch. The tolerance for this subtle structural shift would likely not be seen in crystallographic or bioinformatic studies.

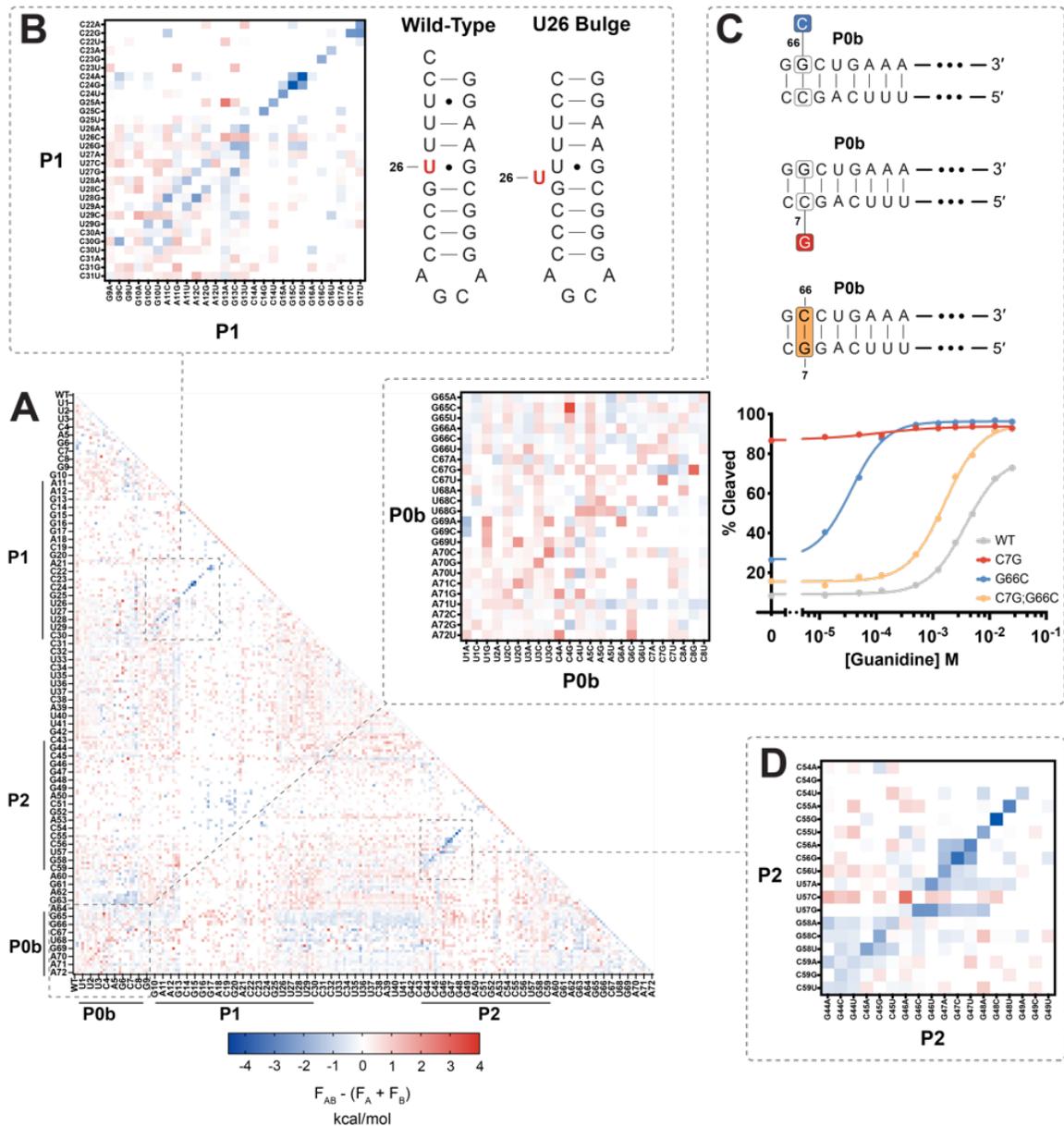


Figure 3.4 Double mutant interactions reveal structural features A) Epistasis heatmap of double mutant function parameters. Negative values reveal P1 and P2 helices, while positive values identify P0b. B) Inset of P1 covariation with parallel diagonal indicating the U26 bulge shown in the accompanying secondary structure C) Inset of P0b covariation and secondary structure of key single mutations to a P0 base pair and the double mutant restoration D) Inset of P2 covariation.

Covariation also exposes the riboswitch's expression platform. We had previously modeled the P0a and P0b helices based upon the complementarity of the putative aSD element in the linker region with the SD sequence and confirmed that mutations to this region are functionally relevant. However, the extent of this helix and its role as the riboswitch's expression platform had not been measured. As modeled, the P0a helix sequesters the majority of the SD sequence with an aSD element. The P0b helix hybridizes the region between the SD and start codon with the tail 5' to P1. Dimerization of P1 and P2 via a kissing loop interaction in the ON state is proposed to open the P0a and P0b helices by physically separating the strands in space (Figure 3.1A). Interestingly, we observe no evidence for functional covariation in the P0a region. Single and double mutations in this region are also extremely well-tolerated (Figure 3.5). Double mutations to the SD sequence itself are largely functional, indicating that ribosome binding occurs efficiently even given poor pairing between the riboswitch and the 16S rRNA. Thus, sequestration of the majority of the SD sequence is dispensable for riboswitch function.

The true expression platform of the *Pae* Gdm-II riboswitch lies in the P0b helix. Covariation in P0b presents as a red diagonal of positive values, indicating interactions unfavorable to formation of the ON state. This helix is least 7 base pairs and harbors the only positions with variants that break the riboswitch ON (Figure 3.4C). Mutations to G65 and G66 appear to produce drastic improvements in riboswitch function. Thus, covariation presents not as rescue amongst broken variants, but as regression back towards wild-type function through the reinforcement of P0b with the complementary mutation. For example, the G66C variant (Figure 3.4C) responds with an average apparent  $K_{1/2}$  ~65 times tighter than wild-type. Across the base pair, the C7G variant

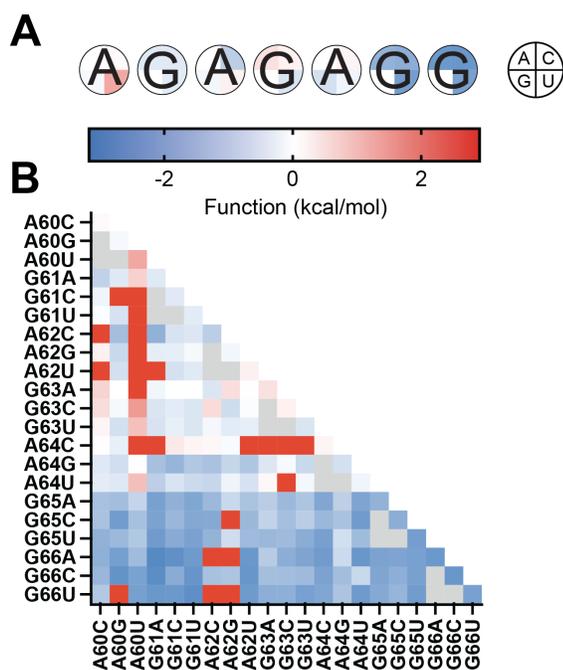


Figure 3.5 Functionality of variants in the SD of the *Pae* Gdm-II riboswitch A) Single mutant function B) Double mutant heat map of SD variants

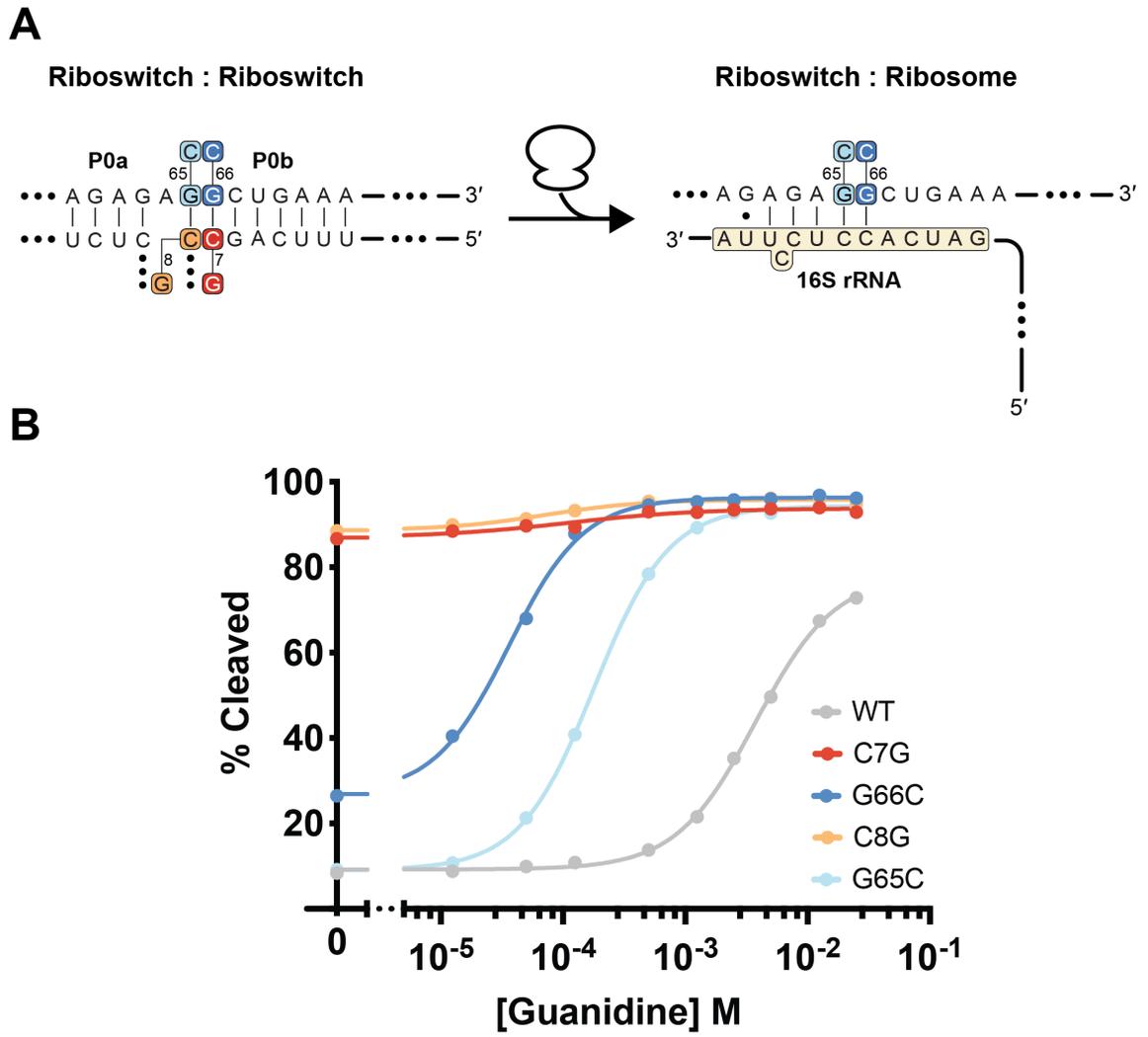


Figure 3.6 P0b mutational asymmetry. A) Secondary structure of the *Pae* Gdm-II riboswitch pairing with its internal anti Shine-Dalgarno compared to pairing with the *E. coli* 16S rRNA anti Shine-Dalgarno B) Representative response curves for transversions to C7, C8, G65 and G66

(Figure 3.4C) is broken ON. The covarying double mutation C7G;G66C (Figure 3.4C) decreases sensitivity back towards the wild-type value (Figure 3.4C).

Interestingly, the increase in sensitivity does not originate simply from the weakening of the P0 helix. Breaking the helix with mutations to C7 and C8 does not produce the dramatic increases in sensitivity seen when breaking the helix with mutations to G65 and G66, the other side of the base pair. Modest increases in functionality are seen with C to U transitions, but transversions completely break the switch ON. Since the 3' side of the helix will interact with the ribosome during initiation, asymmetry across base pairs in the P0b helix is reasonable. Mutations to G65 and G66 weaken P0b, but they also weaken the riboswitch's interaction with the 16s rRNA (Figure 3.6A). These two competing factors result in a riboswitch that is easier to turn ON but is not constitutively ON due to its less effective SD (Figure 3.6B). Mutations to C7 and C8, on the other hand, weaken P0b but maintain strong 16s rRNA pairing. These mutations, therefore, produce a constitutively ON riboswitch (Figure 3.6B).

### **3.2.3 High-throughput Mutational Analysis Confirms Features of Bioinformatic Consensus**

I used these double mutant data to compare prior computational conservation with the biochemical consensus sequence. As previously reported, bioinformatic conservation of the Gdm-II riboswitch class is largely restricted to the two hairpins, P1 and P2 (Figure 3.7A). While the identity and length of the stems of P1 and P2 are not highly conserved, they display a high degree of covariation. The L1 and L2 loops are highly conserved as ACGR tetraloops. The linker region is variable in both length and sequence, and neither the 5' nor 3' tail is conserved.

Dave Hiller and I developed a biochemical consensus diagram of the *Pae* Gdm-II riboswitch based upon the functional data (Figure 3.7B). We screened each single and double mutant based on the function parameter and calculated the conservation of each nucleotide at each position (Materials and Methods). The biochemical conservation of the *Pae* Gdm-II riboswitch recapitulates many features of the bioinformatic consensus. Because bioinformatic conservation reflects features that are maintained across diverse sequences and diverse species, we expect the most highly conserved positions to be immutable in the *Pae* Gdm-II sequence as well. In the biochemical consensus, covariation is observed up to four base pairs in both P1 and P2. The closing base pair of both helices is strongly conserved without covariation, consistent with the structural observation that the closing G-C pairs in both helices engage in  $\pi$ -cation interactions with the ligands (Figure 3.7C).

We also observe strict conservation of the ACGR tetraloops as observed in the bioinformatic consensus. These nucleotides in both loops form the two guanidine binding pockets, and structural studies have shown their direct contact with the ligand (Figure 3.7C). The 3' nucleotide of each loop (A21 and A53 in *Pae* Gdm-II) is bioinformatically conserved as a purine, and we observe this in the biochemical consensus as well. These nucleotides project away from the binding pocket and stack on top of one another in the kissing loop dimerization that occurs upon ligand binding (Figure 3.7D). Since transition from A to G maintains stacking ability, conservation as a purine is functionally reasonable. Previous work with the *Escherichia coli* Gdm-II riboswitch reported an asymmetry in binding between the loops [97]. Mutational analysis indicated a preference for A at the fourth position in both loops, despite having a G endogenously in P2. I also

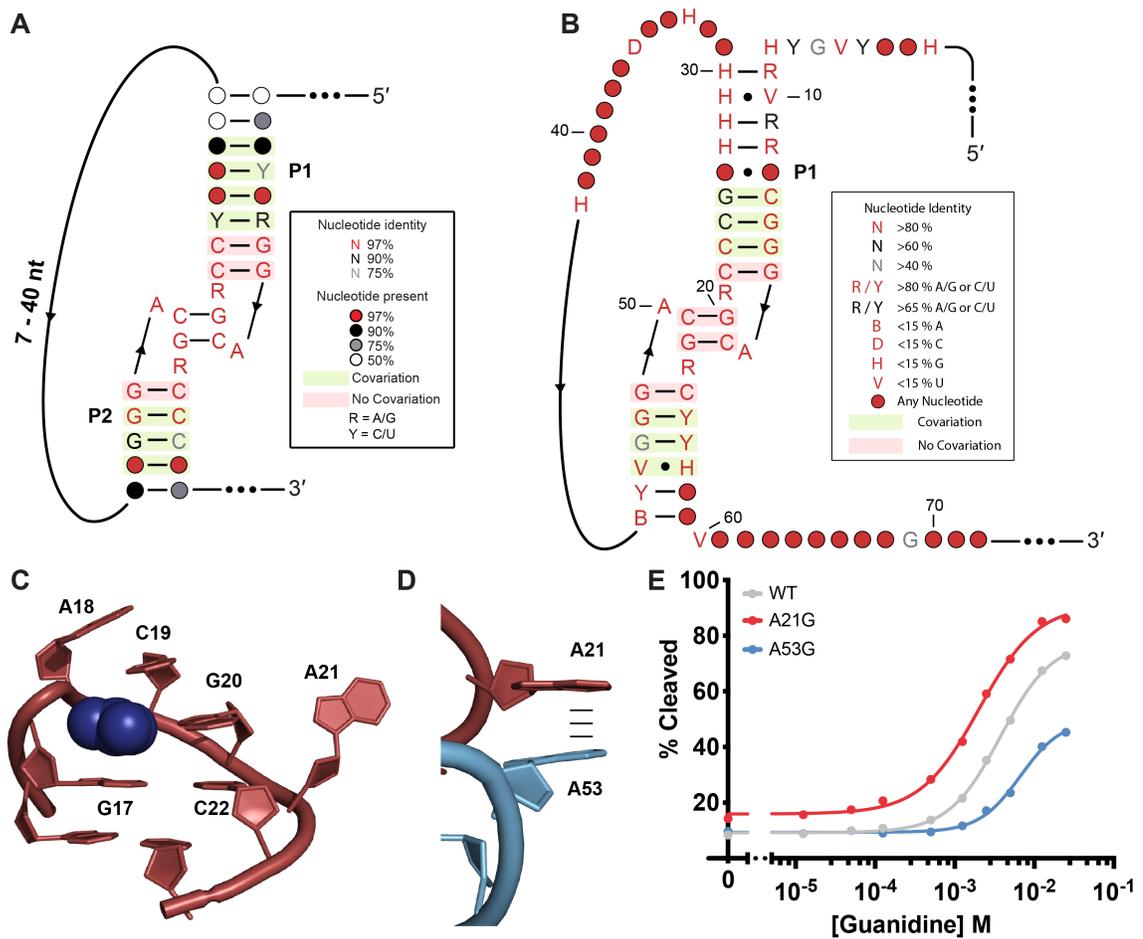


Figure 3.7 Comparison of bioinformatic and biochemical sequence conservation A) Bioinformatic consensus of the Gdm-II riboswitch class B) Biochemical consensus of the *Pae* Gdm-II riboswitch C) Conserved P1 binding pocket nucleotides D) Representation of the stacking interaction between A21 from P1 and A53 from P2. E) Response profiles of two linker nucleotide variants with marked functional deficiencies.

noticed a functional asymmetry between the loops in the *Pae* Gdm-II riboswitch, but while the A to G transition in P2 (A53G) is less functional, the same transition in P1 (A21G) is slightly more functional (Figure 3.7E). Thus, while the bioinformatic consensus indicates an equal tolerance of both A and G at these positions, the individual sequence context of each riboswitch creates a functional preference.

The biochemical consensus has also identified positions that do not appear in the bioinformatic consensus sequence. While mutations to the variable linker region between P1 and P2 have largely no effect on riboswitch function, the U33G or U36C mutation each break the switch into an OFF state (Figure 3.7B). This linker has no published conservation with respect to nucleotide identity or length and appears in none of the available crystal structures. Examining the minimum free energy of the OFF state structures of these mutants illuminates their functional role, however. We hypothesize that the U33G mutation dramatically stabilizes an alternate fold that disrupts the P1 binding site (Figure 3.8A,B). While the SD is not directly sequestered in this competing conformation (Figure 3.8B), the ribosome still struggles to initiate on this sequence at any concentration of guanidine (Figure 3.8C). This mutation is not flagged as detrimental in prior bioinformatic analyses since the alternative conformation is specific to this particular sequence background.

The U36C mutation reveals the delicate energetics of this riboswitch's expression platform. While U36 is part of the linker region that is not bioinformatically conserved, it is biochemically conserved as anything but a C. Interestingly, I identified an interaction between U36C and G6: any mutation to G6 recovered function. I interrogated the connection between these two positions using secondary structure modeling. The

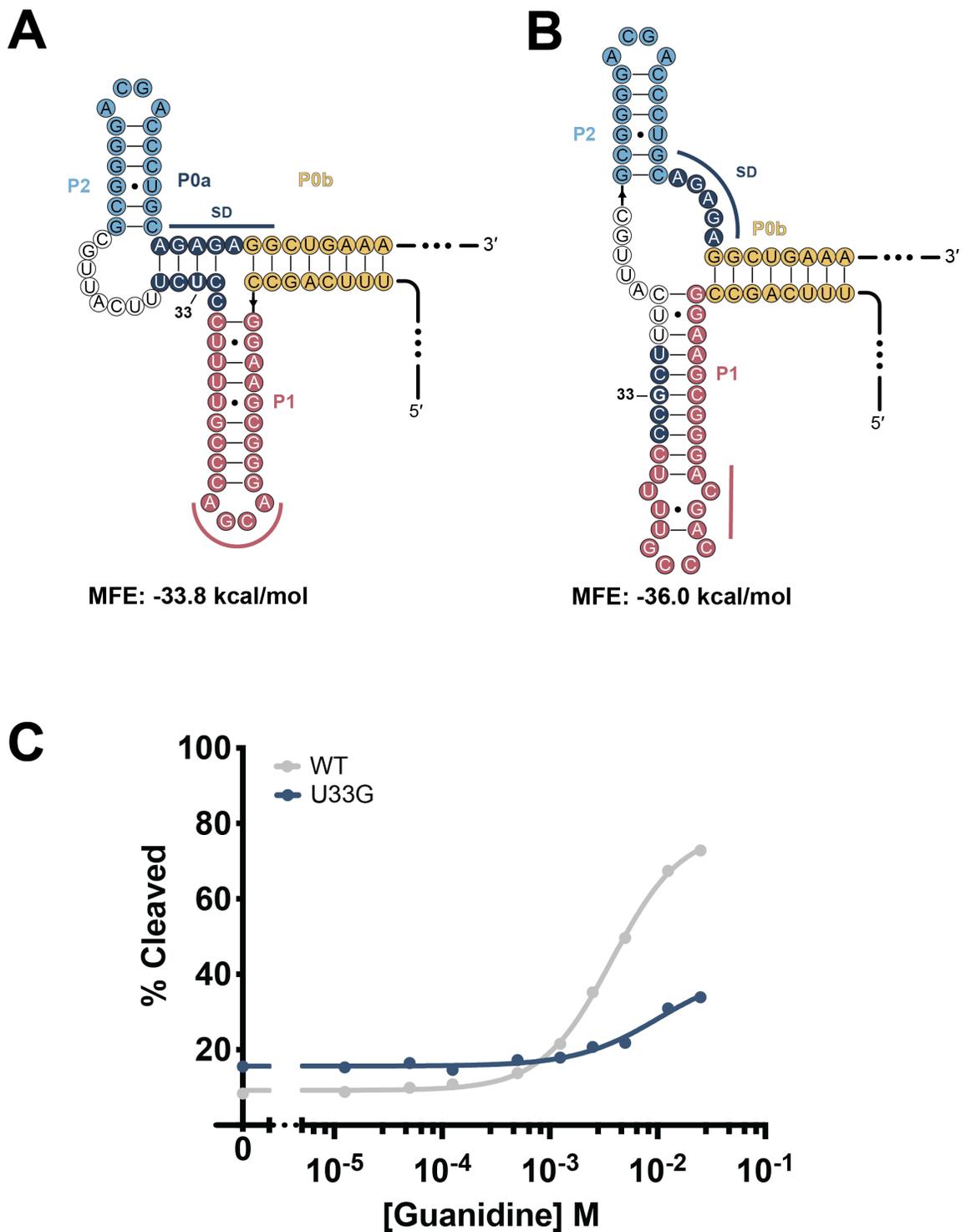


Figure 3.8 Functional deficiency of U33G. A) Modeling the WT MFE OFF state B) Modeling the U33G MFE OFF state C) Representative RelE cleavage guanidinium response profiles of the WT (grey) and U33G (gold) mutant.

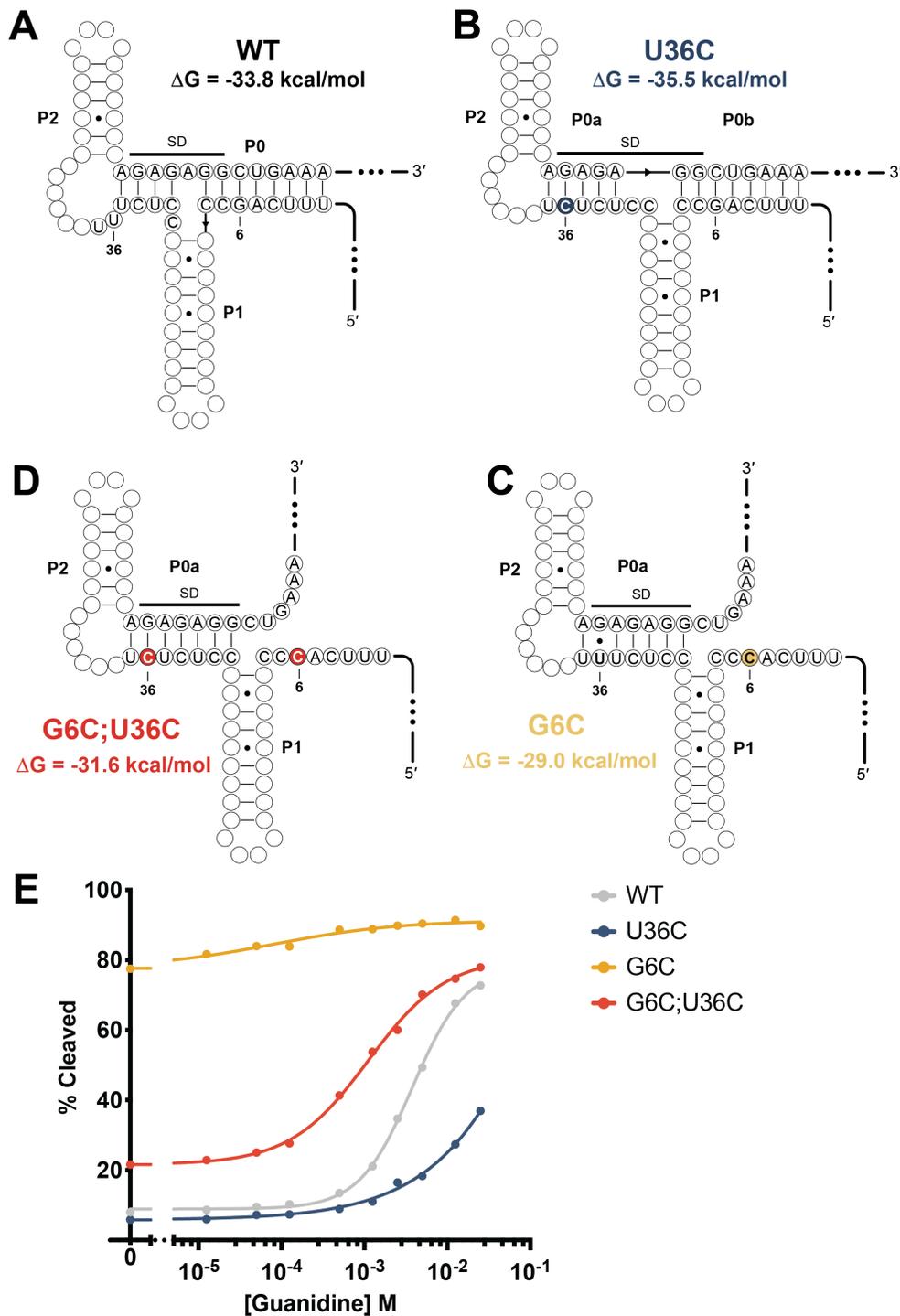


Figure 3.9 Modeling the role of U36C in riboswitch function. A) The WT MFE secondary structure of the *Pae* Gdm-II riboswitch. Ribosome binding site shown in navy. B) The U36C MFE secondary structure. U36C mutation shown in yellow. C) The G6C MFE secondary structure. G6C mutation shown in red. D) The G6C;U36C MFE secondary structure. G6C and U36C mutations shown as light blue. E) Representative response profiles of the WT, G6C, U36C, and G6C;U36C *Pae* Gdm-II riboswitches.

minimum free energy (MFE) structure of the WT sequence includes the previously modeled P0a and P0b helices (Figure 3.9A). P0a sequesters the SD with an aSD element, and P0b comprises the 5' and 3' tails. The U36C mutation produces an even better aSD element, increasing the number of base pairs in P0a to 5 in the MFE (Figure 3.9B). We hypothesize that this OFF state stabilization largely breaks the riboswitch OFF (Figure 3.9E). Mutation of G6 (e.g. G6C), on the other hand, completely opens up the P0b helix in the MFE, expanding P0a to 7 base pairs (Figure 3.9C). Destabilization of the OFF state may be the reason this variant is broken constitutively ON (Figure 3.9C). The G6C;U36C double mutant, however, recovers function with minimal destabilization of the OFF state (Figure 3.9D,E). P0b remains open in this double mutant, but replacing the GU wobble in the 7-base-pair P0a with a GC pair in the MFE may stabilize the OFF state enough to recover ligand responsiveness. Thus, I propose that the U36C point mutation dysregulates riboswitch function in this fixed sequence background by populating an alternate conformation that requires more energy for switching than supplied by ligand binding. Mutations elsewhere in the sequence may energetically compensate and restore the ligand-dependent response.

### **3.2.4 Wobble Base Pairs Affect Apparent Sensitivity and Cooperativity**

While the Hill coefficient produced from fitting the global dataset with the Hill equation is extremely sensitive to the goodness of the fit, some variants display significant differences in the apparent cooperativity of their guanidine response profiles. The wildtype *Pae* Gdm-II riboswitch cooperatively responds to guanidine with a Hill coefficient of  $1.4 \pm 0.2$ , which agrees with the Hill coefficient previously reported for the *G. violaceus* Gdm-II riboswitch [68]. The sensitivity of the Hill coefficient is such that

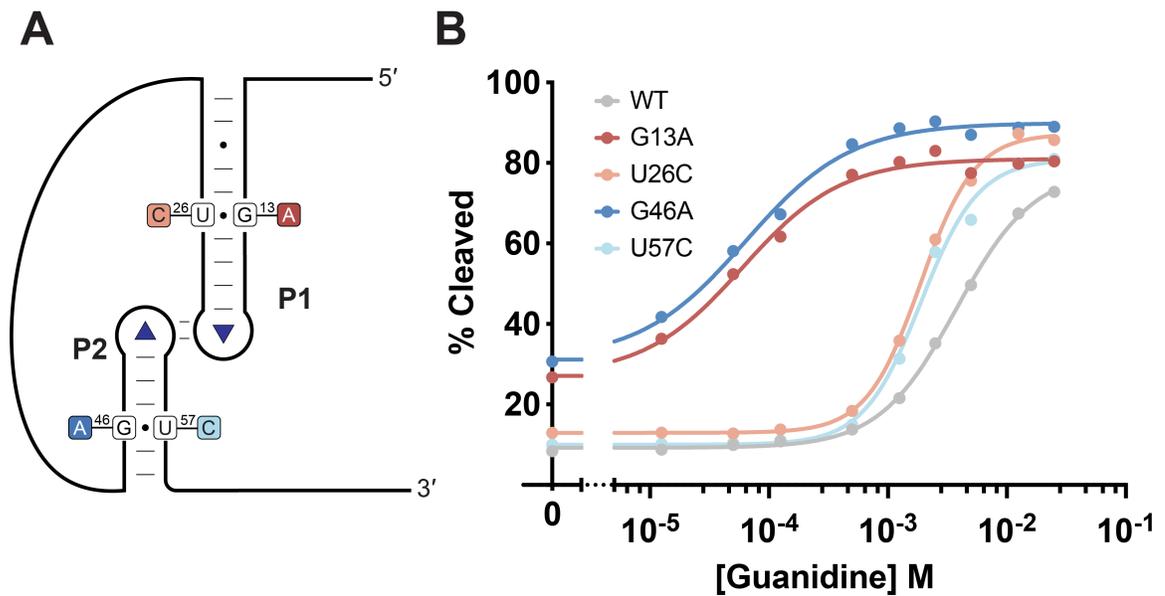


Figure 3.10 A) Secondary structure highlighting the two GU wobbles in the P1 and P2 stems. B) Representative response profiles of the 4 single mutations restoring Watson-Crick pairing in the P1 and P2 wobbles C) Relationship between the average amplitude, sensitivity, and Hill coefficient for the single mutations and covarying double mutations at the P1 and P2 wobbles

we could only reliably calculate Hill coefficients for the single mutants in this dataset, as the double mutants had noisier fits overall. Among the positions with changes in apparent cooperativity, the GU wobble base pairs in P1 and P2 appeared to tune both the sensitivity and the apparent cooperativity of the riboswitch. These base pairs covary in the bioinformatic consensus but are not strictly conserved as GU wobbles. Transitions to canonical G-C base pairs in either P1 or P2 (Figure 3.10A) produce riboswitches that maintain wild-type sensitivity with an average  $K_{1/2}$  of  $1.96 \pm 0.06$  mM (Figure 3.10B). These mutants display heightened positive cooperativity, however, with an average Hill coefficient of  $1.8 \pm 0.2$  (Figure 3.10C). This value is in the fourth quartile of the distribution of Hill coefficient values among functional variants. Transitions to canonical A-U base pairs in either P1 or P2 (Figure 3.10A), however, produce riboswitches with 50-fold tighter  $K_{1/2}$  values on average but with an apparent loss of cooperativity (Figure 3.10B,C). The average Hill coefficient for these A-U base pairs is decreased to  $0.98 \pm 0.2$ , which is in the first quartile. The tightening of sensitivity cannot, therefore, be attributed to stabilization of P1 or P2 since transition to G-C base pairs maintain wild-type cooperativity and sensitivity. These positions are clearly integral to the communication between the two guanidine binding sites, but how mutations at these sites disrupt this energetic network is not immediately apparent.

### 3.3 Discussion

The expansion of the RelE cleavage assay to a sequencing-based readout here establishes an efficient and quantitative approach for the comprehensive mutational analysis of translational riboswitches. We generated quantitative functional data for over 23,000 single and double mutants of the 72-nucleotide *P. aeruginosa* Guanidine-II riboswitch.

Using the functional data generated here, we determined the biochemical conservation of each position in this sequence. Conservation of the binding pockets is consistent with previous bioinformatic analyses, and we hypothesize that variants of the biochemically conserved positions not observed in prior computational and structural studies may stabilize functionally incompetent alternative structures in this fixed sequence background. While there is no evidence for pairing between the SD and aSD in the P0a helix, these data have confirmed the functional relevance of the putative P0b helix and identified key nucleotides that tune the riboswitch's sensitivity and apparent cooperativity.

When riboswitches respond to ligands with poorly understood metabolisms, studying these regulatory systems may also further our understanding of biology. The *Pae* Gdm-II riboswitch cooperatively senses two molecules of the small toxic metabolite guanidine. The riboswitch responds with a weak apparent  $K_{1/2}$  in the low millimolar range. Little is known about guanidine metabolism, although a guanidine hydrolase was recently identified in cyanobacteria [1]. The  $K_M$  of this enzyme is similarly in the low millimolar range, which, in conjunction with the wild-type  $K_{1/2}$  value reported for the *Pae* Gdm-II riboswitch, implies that significant accumulation of guanidine is required to trigger a cellular response. The riboswitch can improve its sensitivity by two orders of magnitude with a single mutation; therefore, we suggest that the riboswitch has been finely tuned to respond to guanidine at a physiologically relevant concentration. Studying riboswitch functional landscapes can facilitate our understanding of previously unknown biology.

Two of the positions that temper sensitivity are located within the predicted P0b helix, outside of the initial bioinformatic consensus and subsequent structural studies. The importance of these positions highlights the value of functional assays to complement bioinformatic identification and binding assays in the development of a wholistic understanding of riboswitch function. For engineering purposes, mutational data factors dramatically into construct development. The single and double mutant fits in this dataset indicate that a wide range of amplitudes and sensitivities are achievable with conservative sequence changes. This comprehensive functional survey of single and double mutations provides an atlas of potential guanidine biosensors covering an extensive range of possible response profiles.

This comprehensive dataset also illuminates the energetics of a translational riboswitch's expression platform for the first time. We observe single mutations to the riboswitch that can break the system ON or OFF. The rescue of G6 variants by U36C indicates that the energetics of the riboswitch expression platform are finely tuned. A single GU to GC transition is the difference between a broken and a functional switch. A similar observation was made for a transcriptional glycine riboswitch [38].

Riboswitches operate within two regimes: kinetic and thermodynamic. Transcriptional riboswitches typically function kinetically since the window for ligand binding is necessarily short [34], [35], [37], [40]. The aptamer must sense the appropriate ligand before the polymerase has transcribed past the terminator. Translational riboswitches, on the other hand, can operate thermodynamically to control gene expression over the lifetime of the transcript [41]. The conformational dynamics of the 5'-UTR populate binding-competent states that can respond to ligand and alter the rate of

ribosome association post-transcriptionally. While a translational guanidine riboswitch and a transcriptional glycine riboswitch are obviously dramatically different elements in sequence, mechanism, and kinetics, the present data indicate that, regardless of regime, the thermodynamics of riboswitch expression platforms are precisely tuned—a single hydrogen bond can make or break them.

These data also indicate that direct pairing of the Shine-Dalgarno sequence is not strictly necessary for controlling gene regulation in this translational riboswitch system. Interestingly, both single and double mutations to the SD sequence have very little effect on translation initiation efficiency, indicating a significant tolerance for deviation from canonical 16S rRNA pairing. We see no evidence of covariation within P0a, which comprises the SD and aSD. The U36C mutant breaks the riboswitch OFF by enhancing the stability of the P0a helix, so P0a can still play a functional role. However, the only variants capable of breaking the riboswitch ON reside in P0b. Pairing in this region exhibits covariation, and structure in between the SD and the start codon is sufficient for effecting ligand-dependent gene regulation. Formation of the P0b helix in the OFF state abrogates ribosome binding and translation initiation. Rather than utilizing the formation of an alternate helix in the ON state, ligand-induced dimerization of the P1 and P2 helices physically separates the two strands of the P0b helix in space to allow for loading of the ribosome on the SD. Thus, the true expression platform, instead, resides in the P0b helix. This finding may assist in the categorization of additional translational riboswitches that do not include obvious anti-SD sequences.

Mutations to the flexible linker region of the *P. aeruginosa* Gdm-II riboswitch highlight variants that are not functionally tolerated in this fixed sequence background.

Bioinformatic consensus diagrams illustrate the conservation of key nucleotides and positions in sequences derived from diverse organisms. These sequences are the product of extensive evolutionary pressure, so conflicting mutations, especially those close in sequence space, would be rapidly eliminated. This comprehensive mutational analysis has experienced no functional selective pressure and instead offers an unbiased view of mutational interactions in this fixed background: both positive and negative interactions are seen within this dataset. The biochemical consensus, therefore, reveals the rugged functional landscape within two mutations of this defined sequence. Mutations to the highly bioinformatically conserved nucleotides are predictably detrimental. The repressive conformational rearrangements we hypothesize are induced by the U33G and U36C mutations indicate that single point mutants outside of the aptamer and expression platform can severely inhibit riboswitch function. Of the possible single mutations to the *Pae* Gdm-II riboswitch, ~50% retain function. Previous work has shown that ~30% of random single amino acid mutations in a given protein were inactivating substitutions [99]. This comparison suggests that within a fixed sequence background, single nucleotide substitutions to a functional RNA may be more detrimental than single amino acid substitutions to a functional protein.

These data highlight the important functional role of the GU wobble base pairs in the P1 and P2 helices. Single mutations restoring canonical Watson-Crick pairings reveal that these positions fine tune the sensitivity and apparent cooperativity of the riboswitch. Explaining the loss of apparent cooperativity with mutations at these positions is more elusive. Cooperativity has been observed in multiple classes of riboswitches including glycine [100], THF [101], cyclic-di-AMP [102] and, more recently, PreQ<sub>1</sub> [48]. Dual

ligand binding may provide the riboswitch with a more sensitive dial, allowing a more digital response to its ligand. Binding two ligands, especially for small metabolites like glycine and guanidine, may provide more energy for conformational switching as well. The loss of apparent cooperativity with a single mutation is, therefore, an intriguing dimension to the *Pae* Gdm-II functional landscape. It could simply reflect a divergence from the energetic assumptions of the Hill equation, including that each binding site has a similar affinity [103]. Cooperative responses are observed when binding to the first site improves the affinity of the second site. This results in a steeper response, since binding a second equivalent is more favorable after the first binding event. However, these mutations to the GU wobble positions may stabilize one binding site to such an extent that the second binding event is no longer significantly more favorable than the first thereby reducing the Hill coefficient back to 1. In a tandem glycine system, remediation of a GU wobble to a GC pair stabilized the weaker aptamer enough to restore cooperativity as evidenced by a Hill coefficient of 1.4 [64]. We may be observing the inverse in this guanidine system, however, where stabilization of one binding site over the other with a GU to GC transition destroys apparent cooperativity by increasing the discrepancy between binding site affinities. Alternatively, these mutations may also indicate a disruption of the communication network between the two binding pockets. Further study is needed to make the distinction.

The use of RelE in high-throughput mutational analysis has allowed us to generate a vast amount of functional data about a translational riboswitch—but this method can be readily applied to other translational regulators. We have previously shown the ability of RelE to report specificity changes in variant riboswitches and quantitative translation

initiation differences in yeast 5'UTR isoforms. Application of this method to RNA thermometers [21], viral IRESes [16], variant riboswitches [104], and T-box RNAs among others will increase our understanding of the biology, evolution, and therapeutic potential of these elements.

## 4 Conclusions & Future Directions

Translation initiation across domains of life is a tightly regulated process. Protein factors and RNA elements temper translation rates both globally and for individual transcripts. Understanding how these regulatory systems function not only enhances our fundamental understanding of biology but also our ability to engineer sensors, circuits, and therapeutics. I have focused on RNA regulators in this thesis, with a special emphasis on translational riboswitches. I have developed a novel method for quantifying translation initiation using the ribosome-dependent endonuclease RelE. I have shown that RelE is sensitive to sequence-driven changes in translation initiation in riboswitches as well as in yeast 5'-UTR isoforms. My extension of this method to a massively parallel mutational analysis of the *P. aeruginosa* Gdm-II riboswitch revealed how the riboswitch's sequence is precisely tuned within its organism to respond with a particular sensitivity and dynamic range. The ease of performing this method with either a gel- or NGS-based readout will hopefully inspire many researchers to use RelE to quantify translation initiation in their system of choice.

This RelE cleavage assay is immediately useful for validating a novel riboswitch's mechanism of function. Given a particular sequence and a particular ligand, we can observe ligand-dependent changes in ribosome association. I have also performed this RelE cleavage assay on the PreQ<sub>1</sub> from *Carnobacterium antarcticus* that has been recently crystallized (Figure 4.1A,B) [48]. This small riboswitch is predicted to function translationally by obscuring the Shine-Dalgarno sequence upon cooperative binding of two molecules of PreQ<sub>1</sub>. While ligand-dependent gene regulation has been observed *in*



*vivo* with fluorescent reporters, the direct control of translation initiation had not been observed. I performed RelE cleavage over a range of PreQ<sub>1</sub> concentrations and observed a cooperative ligand response (Figure 4.1C). The extent of RelE cleavage decreased as the concentration of PreQ<sub>1</sub> increased, consistent with this motif functioning as a translational OFF switch. The  $K_{1/2}$  of this construct is significantly higher than the reported  $K_D$ , which may be an artifact of performing translation initiation at a temperature significantly above the preferred growth temperature of this bacterium. The Hill coefficient of 1.3 indicates that the riboswitch does, in fact, function cooperatively. Thus, I have confirmed that the *Can* PreQ<sub>1</sub> riboswitch functions as a cooperative OFF switch at the level of translation initiation.

This assay is also sensitive to changes in ligand specificity. RelE cleavage can therefore be used to screen a given riboswitch scaffold against a full panel of ligands to easily define the sequence's specificity landscape. Multiple classes of variant riboswitches—classes that use similar architectures to recognize different ligands—have been identified to date [67], [88], [104]–[110]. With the NGS-based readout, diverse mutant libraries of variant riboswitches can be screened against multiple ligands concurrently to probe the functional landscape of a given scaffold. For the cyclic-di-GMP (cdG) riboswitch, for example, a natural variant responds promiscuously to both cdG and cyclic-AMP-GMP (cAG) and functions translationally (Strobel Lab, unpublished). Performing RelE cleavage on a mutant library of this promiscuous riboswitch in the presence of cdG or cAG could identify the sequence determinants for riboswitch specificity—and not simply binding specificity, but functional specificity.

I am especially excited about the prospect of utilizing RelE to examine the functional landscape of viral internal ribosome entry sites (IRESes). RelE has previously been used with IRESes to locate the A-site of the initiated ribosome via a “RelE print” [24], [51], [111]. However, RelE has not been used to quantify translation initiation mediated by these elements. My previous work in wheat germ extract with yeast 5'-UTR isoforms indicates that RelE cleavage can sensitively distinguish between sequences that efficiently initiate and those that initiate poorly. Therefore, given a comprehensive mutant library of a given IRES, RelE will be able to distinguish mutations that sustain, enhance, or destroy the element's ability to bind to ribosomes and initiate translation. This atlas of mutations may reveal important functional interactions within the element, an important step in engineering and therapeutic design efforts.

RelE cleavage provides a precise, quantitative measure of translation initiation. As such, the assay reported here can define sequence-to-function relationships in *cis*-regulatory RNA elements. Beyond the comprehensive mutational analysis of known motifs described above, RelE cleavage can also serve as validation of novel regulatory elements identified through high-throughput techniques. I have leveraged RelE to probe the function of translational riboswitches and uncovered the delicate functional tuning of the sensitivity and cooperativity of the *P. aeruginosa* Gdm-II riboswitch. I look forward seeing what else RelE cleavage can reveal about additional translational regulators.

## 5 Materials and Methods

### 5.1 Materials and Methods: Quantifying Translation

#### Initiation with RelE Cleavage

##### 5.1.1 Design of riboswitch constructs

The wild-type sequence for the *Pseudomonas aeruginosa* Guanidine-II riboswitch is characterized previously [83]. ~20 nucleotides past the endogenous start codon were taken from the *P. aeruginosa* PAO1 genome. The second codon was mutated to a TAG stop codon for efficient RelE cleavage. The wild-type sequence for the *Vibrio vulnificus* adenine-sensing *add* riboswitch was previously reported [43] and adapted for our study. The 3' end was shortened to ~20 nucleotides past the start codon, and the second codon was similarly mutated to a TAG stop codon. The T7 promoter was added to the 5' end of both sequences for *in vitro* transcription. Full sequences and mutant sequences are provided in the Appendix.

##### 5.1.2 RNA preparation and labeling

RNA was transcribed directly off oligonucleotides ordered either from Keck Oligo Synthesis Resource at Yale University or Integrated DNA Technologies (IDT) and purified via denaturing 10% PAGE. For 5' radiolabeling, riboswitch RNA was dephosphorylated with Antarctic Phosphatase (NEB) and labeled with  $^{32}\text{P}$ - $\gamma$ -ATP (Perkin Elmer) via T4 PNK (NEB). For yeast 5'UTR isoforms, RNA was capped with  $^{32}\text{P}$ - $\alpha$ -GTP via the Vaccinia Capping System (NEB). Radiolabeled RNA was similarly purified via denaturing PAGE or by passing through a G-25 spin column (Cytiva).

### 5.1.3 Translation Initiation and RelE cleavage

For riboswitch constructs, RNA refolding solutions contained the following: 100 nM RNA, 1X 219H Buffer (50 mM HEPES-KOH pH 7.5, 70 mM NH<sub>4</sub>Cl, 30 mM KCl, 7 mM MgCl<sub>2</sub>), variable concentrations of ligand, and 5'radiolabeled RNA in trace.

Guanidine riboswitch RNA was refolded by heating to 90°C for 2 minutes, slowly cooled to 37°C at a ramp rate of 0.1°C/s, and then held at room temperature. Adenine riboswitch RNA was refolded by heating to 95°C for 5 minutes and then immediately cooling to 4°C.

Translation initiation solutions contained the following: 1X 219H Buffer, 100 nM IF1, 100 nM IF2, 100 nM IF3, 1 mM GTP, 100 nM fMet-tRNA<sup>fMet</sup>, 100 nM 70S *E. coli* ribosomes, and 10 nM RNA. For the guanidine riboswitch, ligand concentrations ranged between 0 and 25 mM Guanidine Hydrochloride. For the adenine riboswitch, ligand concentrations ranged between 0 and 500 μM Adenine. Translation initiation solutions were incubated at 37°C for either 15 minutes (guanidine riboswitch) or 30 minutes (adenine riboswitch) before being incubated with 1 μM RelE. Cleavage reactions were quenched with the addition of an equal volume of formamide loading buffer (FLB).

For reactions in a commercially available protein expression system, 10 nM refolded *Pae* Gdm-II RNA was incubated in the PURExpress ΔRF123 (NEB) system for 30 min at 37°C before being incubated with 1 μM RelE. 5' radiolabeled RNA was present in trace. Guanidine hydrochloride concentrations ranged between 0 and 25 mM. Cleavage reactions were immediately quenched with the addition of an equal volume of formamide loading buffer (FLB).

For yeast 5'UTR constructs, trace amounts of radiolabeled RNA were incubated with 50% (v/v) Wheat Germ Extract (Promega), 0.8U/μL RNaseOUT (Invitrogen), 120 mM potassium acetate, 80 μM methionine, and 5 μM RelE for 30 minutes at 25°C. Reactions were then quenched with the addition of an equal volume of FLB.

#### **5.1.4 Denaturing PAGE and Analysis**

Radiolabeled products were separated on a 10% polyacrylamide gel, visualized with a Typhoon (GE), and analyzed in ImageQuant (GE). Assuming the concentration of labeled RNA is much less than the  $K_{1/2}$ , the guanidine data were fit in PRISM with the following equation:

$$Y = Amplitude \left[ \frac{X^n}{(K^n + X^n)} \right] + Ymin \quad (1)$$

The adenine data were fit in PRISM with the following equation:

$$Y = Amplitude \left[ \frac{X}{K+X} \right] + YMin \quad (2)$$

For the yeast 5'UTR constructs, constructs were pairwise compared in PRISM using an unpaired two-tailed t test with Welch's correction.

## **5.2 Materials and Methods: Guanidine-II Mutational**

### **Analysis**

#### **5.2.1 Design of riboswitch constructs**

The wild-type sequence for the *Pseudomonas aeruginosa* Gdm-II riboswitch has been characterized previously [83]. For sequencing purposes, a 22 nucleotide handle was added to the 5' end of the riboswitch, and ~20 nucleotides past the endogenous start codon were taken from the *P. aeruginosa* PAO1 genome. The second codon was mutated to a TAG stop codon for efficient RelE cleavage, and GTA was added to the 3' terminus such that the cleaved and full length products contained the same 3' triplet to eliminate potential ligation bias. The mutant library was created via doped oligos, where each mutated position was doped at rate of 4.2% (1.4% for each substituting nucleotide).

#### **5.2.2 RNA preparation and labeling**

RNA was transcribed directly off oligonucleotides ordered from Keck Oligo Synthesis Resource at Yale University as previously described and purified via denaturing PAGE. For 5' radiolabeling, RNA was dephosphorylated with Antarctic Phosphatase (NEB) and labeled with  $^{32}\text{P}$ - $\gamma$ -ATP (Perkin Elmer) via T4 PNK (NEB). Radiolabeled RNA was similarly purified via denaturing PAGE.

#### **5.2.3 Translation Initiation and RelE cleavage**

RNA refolding solutions contained the following: 100 nM RNA, 1X 219H Buffer (50 mM HEPES-KOH pH 7.5, 70 mM  $\text{NH}_4\text{Cl}$ , 30 mM  $\text{KCl}$ , 7 mM  $\text{MgCl}_2$ ), and 0-1.25 M guanidine hydrochloride. For gel-based assays, refolding solutions also contained  $^{32}\text{P}$ -radiolabeled RNA in trace. RNA was refolded in a thermocycler by heating to 90°C for

2 minutes and slow cooling. Translation initiation solutions contained the following: 1X 219H Buffer, 100 nM IF1, 100 nM IF2, 100 nM IF3, 1 mM GTP, 100 nM fMet-tRNA<sup>fMet</sup>, 100 nM 70S *E. coli* ribosomes, 10 nM RNA, and 0-25 mM guanidine hydrochloride. IF1, IF2, and IF3 were purified from *E. coli* as previously described [112]. fMet-tRNA<sup>fMet</sup> was charged with S-100 lysate as previously described [113]. Translation initiation solutions were incubated at 37°C for 15 minutes before being incubated with 1 μM RelE. Cleavage reactions were quenched with the addition of an equal volume of either FLB (25 mM EDTA, <0.1% Xylene Cyanol, <0.1% Bromophenol Blue, ~93% formamide) for gel-based assays or SPRI Binding Solution (2.5 M NaCl, 20% PEG) for sequencing-based assays.

#### **5.2.4 Preparation of RNA for high-throughput sequencing**

An equal volume of SPRI beads (Bulldog Bio) were added to each quenched RelE reaction tube and mixed thoroughly. Bead purification was performed by incubating for 15 minutes, washing the pelleted beads twice with 70% ethanol, and eluting into water. The 2'3' cyclic phosphate left by RelE cleavage was healed with T4 PNK (NEB). The T4 PNK master mix was added to each RNA sample, and the reaction was performed with-bead at 37°C for 1 hour. An equal volume of SPRI Binding Solution was added to each sample, and bead purification was performed as previously described.

A preadenylated DNA adaptor (5rApp/NNNNNCTGTAGGCACCATCAAT/3ddC/; ordered from IDT) was ligated onto the 3' end of the eluted RNA via T4 RNA Ligase II KQ (NEB). The ligation mixture contained 1X T4 Ligase Buffer, 100 μM DNA adaptor, 25% PEG8000, 10U/μL T4 RNA Ligase II KQ and was incubated at room

temperature overnight. An equal volume of SPRI binding solution was added to each sample, and bead purification was performed as previously described.

Using the ligated adaptor as a handle, reverse transcription was used to convert the RNA into cDNA. The RT primer was annealed to the RNA during a 5 min incubation at 65°C in the presence of dNTPs before cooling to 4°C for 2 minutes. The remaining reagents were then added such that the final RT reaction contained 1X SSIV buffer, 500 µM dNTPs, 25 nM primer, 0.2 U/µL RNaseOUT (Invitrogen), 10 U/µL SuperScript IV (Thermo Fisher), and RNA. The reaction was incubated at 55°C for 45 minutes and then heated to 85°C for 5 minutes to heat inactivate the SSIV enzyme. RNaseA (0.5 µg/µL; Thermo Fisher) and RNaseH (0.1 U/µL; Invitrogen) were added, and the mixture was incubated at 37°C for 1 hour to degrade the RNA. The crude RT mixture was then used as the template for 12 rounds of PCR, during which the barcoded Illumina adaptors were added to the 5' and 3' termini. 1.2X SPRI beads were added to the PCR reactions, and the DNA was bead purified as previously described, except the DNA was eluted into 10 mM Tris-HCl, pH 7. The DNA concentrations were measured via Qbit so that the samples could be combined at approximately equal ratios for submission to the Yale Center of Genome Analysis (YCGA) for sequencing.

### **5.2.5 High-throughput sequencing**

Samples from two complete replicates were sequenced at the YCGA on an Illumina HiSeq 4000 (2 X 150 bp). For each replicate, samples were pooled and sequenced using 30% of a lane for (~3-4% of a lane per sample). Demultiplexing was done by the YCGA.

## 5.2.6 Analysis of sequencing results

The 5' constant region and the 3' adaptor were removed from the sequencing reads using CutAdapt [114]. The remaining sequences were aligned to a full-length WT sequence using Bowtie2 [115]. Discordant alignments were not permitted. Custom Python scripts previously reported [38] were used to determine the fraction of cleaved RNA for all variants with 0-2 mutations at each ligand concentration tested. Reads were classified according to the mutations contained through position 72 (the last nucleotide before the start codon). Reads were labeled as cleaved if the last nucleotide was between positions 75-77. Reads were labeled as full-length if the last nucleotide fell after position 85. The number of full-length and cleaved reads were then counted for each variant to determine the fraction of cleaved RNA at each ligand concentration. The data were fit with the Hill Equation in R.

$$Y = \textit{Amplitude} \left[ \frac{X^n}{(K^n + X^n)} \right] + Y_{\min} \quad (1)$$

X is the concentration of ligand; Y is the percent cleaved; K is the  $K_{1/2}$ ;  $Y_{\min}$  is the minimum percent cleaved; amplitude is the difference between  $Y_{\max}$  and  $Y_{\min}$ ; n is the Hill coefficient. Curves were plotted and visualized in Prism 9.

Values were mapped onto secondary structures using custom Python scripts. Double mutant heatmaps were generated in Prism 9.

### 5.2.7 Calculation of function parameter

Noisy fits were excluded if the  $K_{1/2} < K_{1/2\_SD}$  and/or  $Amp < Amp\_SD$ . The apparent  $K_{1/2}$  value was converted into a free energy according to the following equation:

$$\Delta G_{Sensitivity} = -RT \ln (K_{1/2}) \quad (2)$$

The  $\Delta\Delta G_{Sensitivity}$  was calculated by subtracting the WT value from the variant value. The amplitude was transformed into a pseudoenergy according to the following equation:

$$\Delta G_{Amplitude} = -RT \ln \left( \frac{Amp}{100 - Amp} \right) \quad (3)$$

The  $\Delta\Delta G_{Amplitude}$  was similarly calculated by subtracting the WT value from the variant value. The functionality parameter F was calculated from the sum of the two energies as follows:

$$Function = \Delta\Delta G_{Sensitivity} + \Delta\Delta G_{Amplitude} \quad (4)$$

To calculate epistasis, the average function parameter for each individual mutation was calculated ( $F_i$ ). The sum of the individual mutations was then subtracted from the double mutation function:

$$\Delta F = F_{AB} - (F_A + F_B) \quad (5)$$

### 5.2.8 Generating a biochemical consensus

The frequency distribution of function parameters was determined in Prism 9. Sequences were removed as non-functional if the function parameter was one standard deviation above the mean. This generated a pool of functional single mutants and double mutants. The number of times a single mutation (e.g. A21G) appeared in the pool of functional single and double mutants ( $N_{i,j}$ ) was then determined, where i indicates the position (e.g.

A21) and j indicates the mutation (e.g. G) such that i,j identifies the specific variant (e.g. A21G).

For the WT nucleotide at each position, the number of functional single mutants at that position was subtracted from the total number of functional singles:

$$N_{i,WT} = N_{Func}^S - N_{Func}^S(i) \quad (6)$$

The proportion of each nucleotide at each position was calculated by dividing its number of functional variants by the sum of the functional number for all mutations and WT at that position.

$$P = \frac{N_{i,j}}{\sum N_i} \quad (7)$$

Cutoffs for nucleotide identity and purine/pyrimidine distinctions were made by plotting the frequency distribution of nucleotide proportions in Prism and visually assigning cutoffs as shown in Figure 5.2.8.1.

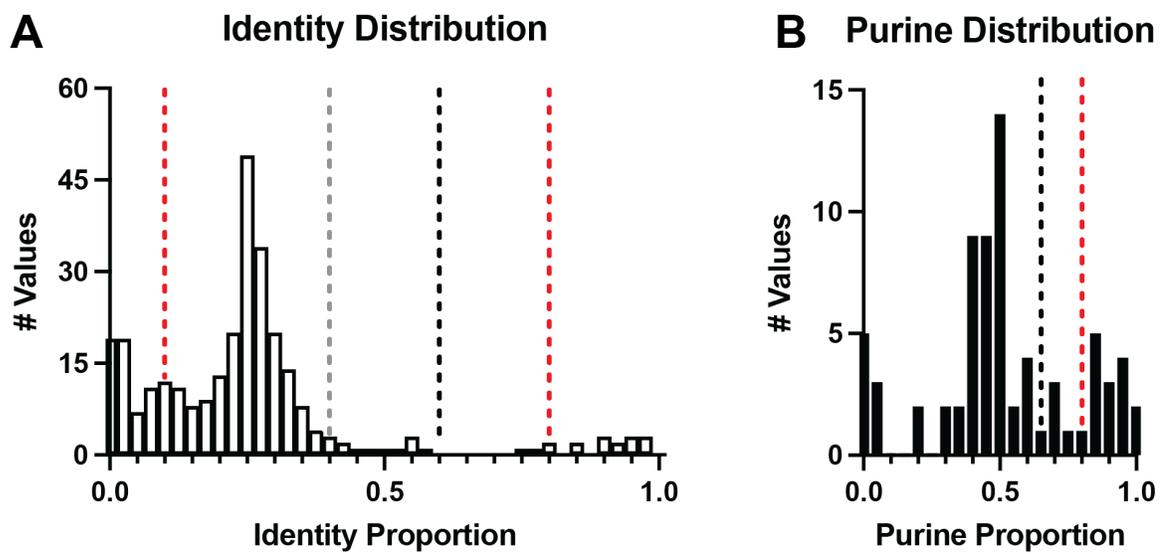


Figure 5.2.8.1 Cutoffs of nucleotide proportion for the biochemical consensus. A) Individual identity proportion distribution.  $>0.8$  is red,  $>0.6$  is black,  $>0.4$  is grey. B) Purine identity distribution.  $>0.8$  is red,  $>0.65$  is black

## 5.2.9 Modeling Minimum Free Energy Conformations

The minimum free energy (MFE) and structure for WT and select mutant constructs were determined by RNAfold [116] using the following command: RNAfold -p -d2 --noLP

## 5.2.10 DNA oligonucleotides and chemicals

All DNA oligos were synthesized by the W.M. Keck Oligonucleotide Synthesis Facility at Yale University unless noted otherwise. Guanidine hydrochloride was obtained from Sigma.

## 5.3 Materials and Methods: *Can* PreQ<sub>1</sub> RelE Cleavage

### 5.3.1 RNA Preparation and Labeling

RNA was ordered from Dharmacon and gel purified by Griffin Schroeder. Purified RNA was labeled with <sup>32</sup>P-γ-ATP (Perkin Elmer) via T4 PNK (NEB). Radiolabeled RNA was purified by passing through a G-25 spin column (Cytiva).

### 5.3.2 RelE Cleavage Assay and Analysis

Translation initiation and RelE cleavage was performed as described above in section 5.1.3. Concentrations of PreQ<sub>1</sub> ranged from 0 – 500 μM, and translation initiation reactions were incubated at 37°C for 30 minutes prior to RelE cleavage. Full-length and cleaved were separated by denaturing PAGE and analyzed as described above in section 5.1.4. Data were fit in Prism 9 with the following equation:

$$Y = \textit{Amplitude} \left[ \frac{X^n}{(K^n + X^n)} \right] + \textit{Ymin} \quad (1)$$

X is the concentration of ligand; Y is the percent cleaved; K is the  $K_{1/2}$ ;  $Y_{\min}$  is the minimum percent cleaved; amplitude is the difference between  $Y_{\max}$  and  $Y_{\min}$ ; n is the Hill coefficient.

## 6 Appendix

### 6.1 RNA sequences for gel-based RelE assay

#### *P. aeruginosa* Gdm-II WT

5'-UUUCAGCCGGAAGCGGGACGACCCGUUUUCCCUCUUUCAUUGCGCGGGG  
ACGACCCUGCAGAGAGGCUGAAAAUGUAGUGGAUCUAUCUCUUGCUCGUA

#### *P. aeruginosa* Gdm-II ON

5'-UUUCAGAAAGGAAGCGGGACGACCCGUUUUCCAAUUUCAUUGCGCGGGG  
ACGACCCUGCAGAGAGGCUGAAAAUGUAGUGGAUCUAUCUCUUGCUCGUA

#### *P. aeruginosa* Gdm-II OFF

5'-UUUCAGCCGGAAGCGGGAAACCCGUUUUCCCUCUUUCAUUGCGCGGGG  
ACGACCCUGCAGAGAGGCUGAAAAUGUAGUGGAUCUAUCUCUUGCUCGUA

#### *V. vulnificus add*

5'-GCUUCAUAUAAUCCUAAUGAU AUGGUUUGGGAGUUUCUACCAAGAGCC  
UUAACUCUUGAUUAUGAAGUCUGUCGCUUUAUCCGAAAUUUUAUAAAGA  
GAAGACUCAUGUAGAAUUACUUUGACCUGCCG

#### *V. vulnificus add* U61C

5'-GCUUCAUAUAAUCCUAAUGAU AUGGUUUGGGAGUUUCUACCAAGAGCC  
UUAACUCUUGAUAUGAAGUCUGUCGCUUUAUCCGAAAUUUUAUAAAGA  
GAAGACUCAUGUAGAAUUACUUUGACCUGCCG

#### YGR196C Short Isoform

5'-GGUUUGAUCAUUACCCUUUUUCUGGAAAGCGGAUAUUUAUUAUGUAGA  
CUGAACAGGUCGGUAGAAAGAAA

#### YGR196C Long Isoform

5'-GAUUAUAGUUAAAAGGGACUGUUUGAUCAUUACCCUUUUUCUGGAAAG  
CGGAUAUUUAUUAUGUAGACUGAACAGGUCGGUAGAAAGAAA

#### YGR196C Long Isoform No Enhancer

5'-GAUUCCAGUUCCCCGGGACUGUUUGAUCAUUACCCUUUUUCUGGAAAG  
CGGAUAUUUAUUAUGUAGACUGAACAGGUCGGUAGAAAGAAA

#### YML069W Short Isoform

5'-GACAAGUGAAGGAACAAUCUAGUAUUGUUGAACAAAGAAUUAUUAUGUAGA  
CCGACUUUGAUAGAAUUUACUUGAA

#### YML069W Long Isoform

5'-GAGAAGGAAAUACCGUUAACAAGUGAAGGAACAAUCUAGUAUUGUUGA  
ACAAGAAUUAUUAUGUAGACCGACUUUGAUAGAAUUUACUUGAA

## 6.2 DNA sequences used in NGS-based RelE assay

### *P. aeruginosa* Gdm-II Template

```
TAATACGACTCACTATAGCGTTATCTCTTATCAGGCAGGTTTCAGCCGGAAGC  
GGGACGACCCGTTTTCCCTCTTTCATTGCGCGGGGACGACCCTGCAGAGAGG  
CTGAAAATGTAGTGGATCTATCTCTTGCTCGTA
```

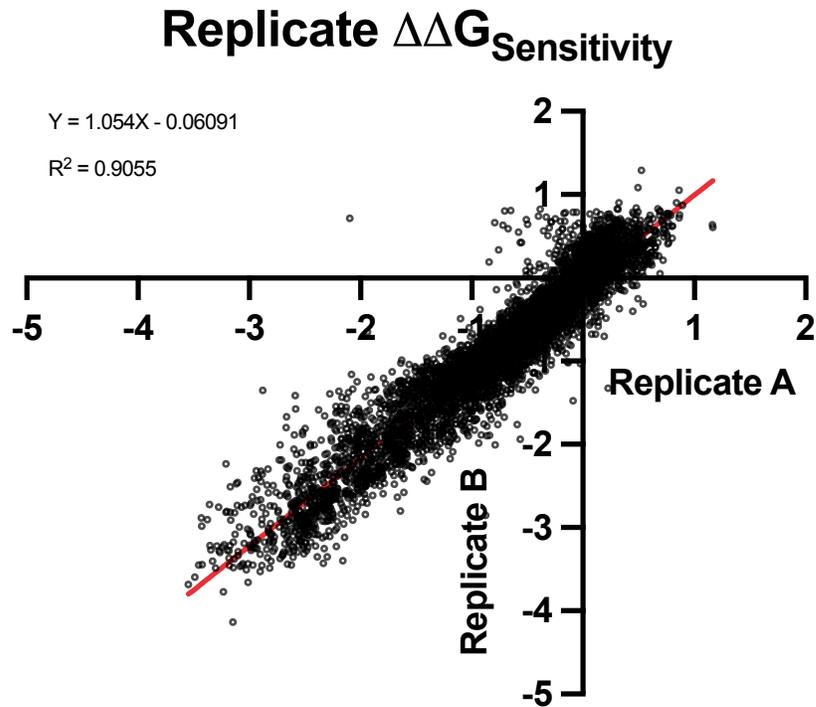
### 3' Ligated Adaptor

```
5'-rApp-NNNNNCTGTAGGCACCATCAAT-ddC
```

### RT Primer

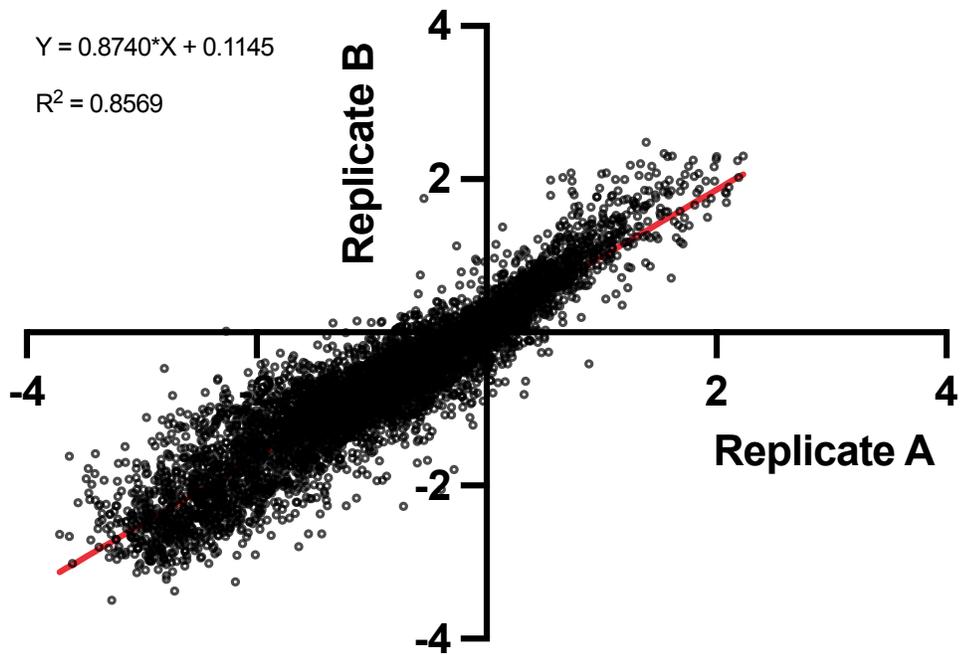
```
5'-GATTGATGGTGCCTACAG
```

## 6.3 NGS Replicate Correlations



Appendix Figure 6.3.1. Correlation of single and double mutant  $\Delta\Delta G_{\text{Sensitivity}}$  between two replicates

# Functionality



Appendix Figure 6.3.2 Correlation of single and double mutant function parameter between two replicates

## 6.4 RNA sequence of *Can* PreQ<sub>1</sub> riboswitch

5'-UGUGGUUCGCAACCAUCCACAUAAAAAAAAACUAGGAGGAUUCACACAUG  
UAGACAAAAAAAAUAUCGUCACAAA-3'

## References

- [1] O. T. Avery, C. M. MacLeod, and M. McCarty, “Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal types : Induction of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III,” *J. Exp. Med.*, vol. 79, no. 2, pp. 137–158, Feb. 1944, doi: 10.1084/jem.79.2.137.
- [2] F. Crick, “Central Dogma of Molecular Biology,” *Nature*, vol. 227, no. 5258, Art. no. 5258, Aug. 1970, doi: 10.1038/227561a0.
- [3] M. B. Hoagland, M. L. Stephenson, J. F. Scott, L. I. Hecht, and P. C. Zamecnik, “A Soluble Ribonucleic Acid Intermediate in Protein Synthesis,” *J. Biol. Chem.*, vol. 231, no. 1, pp. 241–257, Mar. 1958, doi: 10.1016/S0021-9258(19)77302-5.
- [4] K. Kruger, P. J. Grabowski, A. J. Zaug, J. Sands, D. E. Gottschling, and T. R. Cech, “Self-Splicing RNA: Autoexcision and autocyclization of the ribosomal RNA intervening sequence of tetrahymena,” *Cell*, vol. 31, no. 1, pp. 147–157, Nov. 1982, doi: 10.1016/0092-8674(82)90414-7.
- [5] C. Guerrier-Takada, K. Gardiner, T. Marsh, N. Pace, and S. Altman, “The RNA Moiety of Ribonuclease P is the Catalytic Subunit of the Enzyme,” *Cell*, vol. 35, no. 3, Part 2, pp. 849–857, Dec. 1983, doi: 10.1016/0092-8674(83)90117-4.
- [6] W. Gilbert, “Origin of Life: The RNA world,” *Nature*, vol. 319, no. 6055, p. 618, Feb. 1986, doi: 10.1038/319618a0.
- [7] T. Godefroy-Colburn, A. D. Wolfe, J. Dondon, M. Grunberg-Manago, P. Dessen, and D. Pantaloni, “Light-Scattering Studies Showing the Effect of Initiation Factors

- on the Reversible Dissociation of *Escherichia coli* Ribosomes,” *J. Mol. Biol.*, vol. 94, no. 3, pp. 461–478, May 1975, doi: 10.1016/0022-2836(75)90215-6.
- [8] J. A. Steitz and K. Jakes, “How Ribosomes Select Initiator Regions in mRNA: Base pair formation between the 3′ terminus of 16S rRNA and the mRNA during initiation of protein synthesis in *Escherichia coli*,” *Proc. Natl. Acad. Sci.*, vol. 72, no. 12, pp. 4734–4738, Dec. 1975, doi: 10.1073/pnas.72.12.4734.
- [9] S. Nakagawa, Y. Niimura, K. Miura, and T. Gojobori, “Dynamic Evolution of Translation Initiation Mechanisms in Prokaryotes,” *Proc. Natl. Acad. Sci.*, vol. 107, no. 14, pp. 6382–6387, Apr. 2010, doi: 10.1073/pnas.1002036107.
- [10] H. J. Beck and I. Moll, “Leaderless mRNAs in the Spotlight: Ancient but Not Outdated!,” *Microbiol. Spectr.*, vol. 6, no. 4, p. 6.4.02, Jul. 2018, doi: 10.1128/microbiolspec.RWR-0016-2017.
- [11] I. V. Boni, D. M. Isaeva, M. L. Musychenko, and N. V. Tzareva, “Ribosome-Messenger Recognition: mRNA target sites for ribosomal protein S1,” *Nucleic Acids Res.*, vol. 19, no. 1, pp. 155–162, Jan. 1991, doi: 10.1093/nar/19.1.155.
- [12] R. M. Sundari, E. A. Stringer, L. H. Schulman, and U. Maitra, “Interaction of Bacterial Initiation Factor 2 with Initiator tRNA,” *J. Biol. Chem.*, vol. 251, no. 11, pp. 3338–3345, Jun. 1976, doi: 10.1016/S0021-9258(17)33442-7.
- [13] R. J. Jackson, C. U. T. Hellen, and T. V. Pestova, “The Mechanism of Eukaryotic Translation Initiation and Principles of Its Regulation,” *Nat. Rev. Mol. Cell Biol.*, vol. 11, no. 2, Art. no. 2, Feb. 2010, doi: 10.1038/nrm2838.

- [14] A. G. Hinnebusch and J. R. Lorsch, “The Mechanism of Eukaryotic Translation Initiation: New Insights and Challenges,” *Cold Spring Harb. Perspect. Biol.*, vol. 4, no. 10, p. a011544, Oct. 2012, doi: 10.1101/cshperspect.a011544.
- [15] M. Kozak, “Point Mutations Define a Sequence Flanking the AUG Initiator Codon that Modulates Translation by Eukaryotic Ribosomes,” *Cell*, vol. 44, no. 2, pp. 283–292, Jan. 1986, doi: 10.1016/0092-8674(86)90762-2.
- [16] Z. A. Jaafar and J. S. Kieft, “Viral RNA Structure-Based Strategies to Manipulate Translation,” *Nat. Rev. Microbiol.*, vol. 17, no. 2, Art. no. 2, Feb. 2019, doi: 10.1038/s41579-018-0117-x.
- [17] P. J. Schlx and D. J. Worhunsy, “Translational Repression Mechanisms in Prokaryotes,” *Mol. Microbiol.*, vol. 48, no. 5, pp. 1157–1169, 2003, doi: 10.1046/j.1365-2958.2003.03517.x.
- [18] C. Chiaruttini and M. Guillier, “On the Role of mRNA Secondary Structure in Bacterial Translation,” *WIREs RNA*, vol. 11, no. 3, p. e1579, 2020, doi: 10.1002/wrna.1579.
- [19] A. Bernardi and P.-F. Spahr, “Nucleotide Sequence at the Binding Site for Coat Protein on RNA of Bacteriophage R17,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 69, no. 10, pp. 3033–3037, Oct. 1972.
- [20] C. Philippe, F. Eyermann, L. Bénard, C. Portier, B. Ehresmann, and C. Ehresmann, “Ribosomal protein S15 from *Escherichia coli* Modulates Its Own Translation by Trapping the Ribosome on the mRNA Initiation Loading Site,” *Proc. Natl. Acad. Sci.*, vol. 90, no. 10, pp. 4394–4398, May 1993, doi: 10.1073/pnas.90.10.4394.

- [21] J. Kortmann and F. Narberhaus, “Bacterial RNA Thermometers: Molecular zippers and switches,” *Nat. Rev. Microbiol.*, vol. 10, no. 4, Art. no. 4, Apr. 2012, doi: 10.1038/nrmicro2730.
- [22] E. Loh, F. Righetti, H. Eichner, C. Twittenhoff, and F. Narberhaus, “RNA Thermometers in Bacterial Pathogens,” *Microbiol. Spectr.*, vol. 6, no. 2, p. 6.2.13, Apr. 2018, doi: 10.1128/microbiolspec.RWR-0012-2017.
- [23] J. Mailliot and F. Martin, “Viral Internal Ribosomal Entry Sites: Four classes for one goal,” *WIREs RNA*, vol. 9, no. 2, p. e1458, 2018, doi: 10.1002/wrna.1458.
- [24] I. S. Abaeva *et al.*, “The Halastavi árva Virus Intergenic Region IRES Promotes Translation by the Simplest Possible Initiation Mechanism,” *Cell Rep.*, vol. 33, no. 10, p. 108476, Dec. 2020, doi: 10.1016/j.celrep.2020.108476.
- [25] R. R. Breaker, “The Biochemical Landscape of Riboswitch Ligands,” *Biochemistry*, vol. 61, no. 3, pp. 137–149, Feb. 2022, doi: 10.1021/acs.biochem.1c00765.
- [26] A. Serganov and E. Nudler, “A Decade of Riboswitches,” *Cell*, vol. 152, no. 1–2, pp. 17–24, Jan. 2013, doi: 10.1016/j.cell.2012.12.024.
- [27] R. R. Breaker, “Riboswitches and Translation Control,” *Cold Spring Harb. Perspect. Biol.*, vol. 10, no. 11, p. a032797, Nov. 2018, doi: 10.1101/cshperspect.a032797.
- [28] C. Husser, N. Dentz, and M. Ryckelynck, “Structure-Switching RNAs: From Gene Expression Regulation to Small Molecule Detection,” *Small Struct.*, vol. 2, no. 4, p. 2000132, 2021, doi: 10.1002/sstr.202000132.

- [29] R. Byrne, J. G. Levin, H. A. Bladen, and M. W. Nirenberg, “The *in vitro* Formation of a DNA-Ribosome Complex,” *Proc. Natl. Acad. Sci.*, vol. 52, no. 1, pp. 140–148, Jul. 1964, doi: 10.1073/pnas.52.1.140.
- [30] M. Mandal and R. R. Breaker, “Gene Regulation by Riboswitches,” *Nat. Rev. Mol. Cell Biol.*, vol. 5, no. 6, pp. 451–463, Jun. 2004, doi: 10.1038/nrm1403.
- [31] K. Hollands *et al.*, “Riboswitch Control of Rho-Dependent Transcription Termination,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 14, pp. 5376–5381, Apr. 2012, doi: 10.1073/pnas.1112211109.
- [32] J. M. Peters, A. D. Vangeloff, and R. Landick, “Bacterial Transcription Terminators: The RNA 3'-End Chronicles,” *J. Mol. Biol.*, vol. 412, no. 5, pp. 793–813, Oct. 2011, doi: 10.1016/j.jmb.2011.03.036.
- [33] I. Gusarov and E. Nudler, “The Mechanism of Intrinsic Transcription Termination,” *Mol. Cell*, vol. 3, no. 4, pp. 495–504, Apr. 1999, doi: 10.1016/S1097-2765(00)80477-3.
- [34] J. K. Wickiser, W. C. Winkler, R. R. Breaker, and D. M. Crothers, “The Speed of RNA Transcription and Metabolite Binding Kinetics Operate an FMN Riboswitch,” *Mol. Cell*, vol. 18, no. 1, pp. 49–60, Apr. 2005, doi: 10.1016/j.molcel.2005.02.032.
- [35] H. Steinert *et al.*, “Pausing Guides RNA Folding to Populate Transiently Stable RNA Structures for Riboswitch-Based Transcription Regulation,” *eLife*, vol. 6, p. e21297, May 2017, doi: 10.7554/eLife.21297.
- [36] G. A. Perdrizet, I. Artsimovitch, R. Furman, T. R. Sosnick, and T. Pan, “Transcriptional Pausing Coordinates Folding of the Aptamer Domain and the

- Expression Platform of a Riboswitch,” *Proc. Natl. Acad. Sci.*, vol. 109, no. 9, pp. 3323–3328, Feb. 2012, doi: 10.1073/pnas.1113086109.
- [37] C. E. Scull, S. S. Dandpat, R. A. Romero, and N. G. Walter, “Transcriptional Riboswitches Integrate Timescales for Bacterial Gene Expression Control,” *Front. Mol. Biosci.*, vol. 7, 2021, Accessed: Jun. 22, 2022. [Online]. Available: <https://www.frontiersin.org/article/10.3389/fmolb.2020.607158>
- [38] C. D. Torgerson, D. A. Hiller, S. Stav, and S. A. Strobel, “Gene Regulation by a Glycine Riboswitch Singlet Uses a Finely Tuned Energetic Landscape for Helical Switching,” *RNA*, vol. 24, no. 12, pp. 1813–1827, Dec. 2018, doi: 10.1261/rna.067884.118.
- [39] W. Winkler, A. Nahvi, and R. R. Breaker, “Thiamine Derivatives Bind Messenger RNAs Directly to Regulate Bacterial Gene Expression,” *Nature*, vol. 419, no. 6910, pp. 952–956, Oct. 2002, doi: 10.1038/nature01145.
- [40] G. Quarta, K. Sin, and T. Schlick, “Dynamic Energy Landscapes of Riboswitches Help Interpret Conformational Rearrangements and Function,” *PLOS Comput. Biol.*, vol. 8, no. 2, p. e1002368, Feb. 2012, doi: 10.1371/journal.pcbi.1002368.
- [41] J.-F. Lemay *et al.*, “Comparative Study between Transcriptionally- and Translationally-Acting Adenine Riboswitches Reveals Key Differences in Riboswitch Regulatory Mechanisms,” *PLoS Genet.*, vol. 7, no. 1, p. e1001278, Jan. 2011, doi: 10.1371/journal.pgen.1001278.
- [42] K. A. Rotstan *et al.*, “Regulation of mRNA Translation by a Photriboswitch,” *eLife*, vol. 9, p. e51737, Feb. 2020, doi: 10.7554/eLife.51737.

- [43] V. de Jesus *et al.*, “Switching at the Ribosome: riboswitches need rProteins as modulators to regulate translation,” *Nat. Commun.*, vol. 12, no. 1, p. 4723, Aug. 2021, doi: 10.1038/s41467-021-25024-5.
- [44] S. Grill, C. O. Gualerzi, P. Londei, and U. Blasi, “Selective Stimulation of Translation of Leaderless mRNA by Initiation Factor 2: Evolutionary implications for translation,” *EMBO J.*, vol. 19, no. 15, pp. 4101–4110, Aug. 2000, doi: 10.1093/emboj/19.15.4101.
- [45] T. Mašek, L. Valášek, and M. Pospíšek, “Polysome Analysis and RNA Purification from Sucrose Gradients,” in *RNA: Methods and Protocols*, H. Nielsen, Ed. Totowa, NJ: Humana Press, 2011, pp. 293–309. doi: 10.1007/978-1-59745-248-9\_20.
- [46] H. A. King and A. P. Gerber, “Translatome Profiling: Methods for genome-scale analysis of mRNA translation,” *Brief. Funct. Genomics*, vol. 15, no. 1, pp. 22–31, Jan. 2016, doi: 10.1093/bfgp/elu045.
- [47] K. Y. Paek *et al.*, “Translation Initiation Mediated by RNA Looping,” *Proc. Natl. Acad. Sci.*, vol. 112, no. 4, pp. 1041–1046, Jan. 2015, doi: 10.1073/pnas.1416883112.
- [48] G. M. Schroeder, C. E. Cavender, M. E. Blau, J. L. Jenkins, D. H. Mathews, and J. E. Wedekind, “A Small RNA that Cooperatively Senses Two Stacked Metabolites in One Pocket for Gene Control,” *Nat. Commun.*, vol. 13, no. 1, Art. no. 1, Jan. 2022, doi: 10.1038/s41467-021-27790-8.
- [49] M. Kozak, “Primer extension Analysis of Eukaryotic Ribosome-mRNA Complexes,” *Nucleic Acids Res.*, vol. 26, no. 21, pp. 4853–4859, Nov. 1998, doi: 10.1093/nar/26.21.4853.

- [50] N. Ontiveros-Palacios, A. M. Smith, F. J. Grundy, M. Soberon, T. M. Henkin, and J. Miranda-Ríos, “Molecular Basis of Gene Regulation by the THI-box Riboswitch,” *Mol. Microbiol.*, vol. 67, no. 4, pp. 793–803, 2008, doi: 10.1111/j.1365-2958.2007.06088.x.
- [51] D. Andreev, V. Hauryliuk, I. Terenin, S. Dmitriev, M. Ehrenberg, and I. Shatsky, “The Bacterial Toxin RelE Induces Specific mRNA Cleavage in the A Site of the Eukaryote Ribosome,” *RNA*, vol. 14, no. 2, pp. 233–239, Feb. 2008, doi: 10.1261/rna.693208.
- [52] M. Gotfredsen and K. Gerdes, “The *Escherichia coli relBE* Genes Belong to a New Toxin–Antitoxin Gene Family,” *Mol. Microbiol.*, vol. 29, no. 4, pp. 1065–1076, 1998, doi: 10.1046/j.1365-2958.1998.00993.x.
- [53] S. K. Christensen, M. Mikkelsen, K. Pedersen, and K. Gerdes, “RelE, a Global Inhibitor of Translation, is Activated During Nutritional Stress,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 98, no. 25, pp. 14328–14333, Dec. 2001, doi: 10.1073/pnas.251327898.
- [54] K. Pedersen, A. V. Zavialov, M. Yu. Pavlov, J. Elf, K. Gerdes, and M. Ehrenberg, “The Bacterial Toxin RelE Displays Codon-Specific Cleavage of mRNAs in the Ribosomal A Site,” *Cell*, vol. 112, no. 1, pp. 131–140, Jan. 2003, doi: 10.1016/S0092-8674(02)01248-5.
- [55] J. M. Hurley, J. W. Cruz, M. Ouyang, and N. A. Woychik, “Bacterial Toxin RelE Mediates Frequent Codon-independent mRNA Cleavage from the 5′ End of Coding Regions in Vivo\*,” *J. Biol. Chem.*, vol. 286, no. 17, pp. 14770–14778, Apr. 2011, doi: 10.1074/jbc.M110.108969.

- [56] K. Gerdes, “Toxin-Antitoxin Modules May Regulate Synthesis of Macromolecules during Nutritional Stress,” *J. Bacteriol.*, vol. 182, no. 3, pp. 561–572, Feb. 2000, doi: 10.1128/JB.182.3.561-572.2000.
- [57] J.-Y. Hwang and A. R. Buskirk, “A Ribosome Profiling Study of mRNA Cleavage by the Endonuclease RelE,” *Nucleic Acids Res.*, vol. 45, no. 1, pp. 327–336, Jan. 2017, doi: 10.1093/nar/gkw944.
- [58] C. Tuerk and L. Gold, “Systematic Evolution of Ligands by Exponential Enrichment: RNA ligands to bacteriophage T4 DNA polymerase,” *Science*, vol. 249, no. 4968, pp. 505–510, Aug. 1990, doi: 10.1126/science.2200121.
- [59] A. D. Ellington and J. W. Szostak, “*In vitro* Selection of RNA Molecules that Bind Specific Ligands,” *Nature*, vol. 346, no. 6287, Art. no. 6287, Aug. 1990, doi: 10.1038/346818a0.
- [60] J. N. Pitt and A. R. Ferré-D’Amaré, “Rapid Construction of Empirical RNA Fitness Landscapes,” *Science*, vol. 330, no. 6002, pp. 376–379, Oct. 2010, doi: 10.1126/science.1192001.
- [61] J. D. Buenrostro *et al.*, “Quantitative Analysis of RNA-Protein Interactions on a Massively Parallel Array Reveals Biophysical and Evolutionary Landscapes,” *Nat. Biotechnol.*, vol. 32, no. 6, Art. no. 6, Jun. 2014, doi: 10.1038/nbt.2880.
- [62] S. Kobori, Y. Nomura, A. Miu, and Y. Yokobayashi, “High-Throughput assay and Engineering of Self-Cleaving Ribozymes by Sequencing,” *Nucleic Acids Res.*, vol. 43, no. 13, pp. e85–e85, Jul. 2015, doi: 10.1093/nar/gkv265.

- [63] S. Kobori and Y. Yokobayashi, “High-Throughput Mutational Analysis of a Twister Ribozyme,” *Angew. Chem. Int. Ed.*, p. n/a-n/a, Aug. 2016, doi: 10.1002/anie.201605470.
- [64] C. D. Torgerson, D. A. Hiller, and S. A. Strobel, “The Asymmetry and Cooperativity of Tandem Glycine Riboswitch Aptamers,” *RNA*, vol. 26, no. 5, pp. 564–580, May 2020, doi: 10.1261/rna.073577.119.
- [65] J. O. L. Andreasson, A. Savinov, S. M. Block, and W. J. Greenleaf, “Comprehensive Sequence-to-Function Mapping of Cofactor-Dependent RNA Catalysis in the *glmS* Ribozyme,” *Nat. Commun.*, vol. 11, Apr. 2020, doi: 10.1038/s41467-020-15540-1.
- [66] J. L. Baker, N. Sudarsan, Z. Weinberg, A. Roth, R. B. Stockbridge, and R. R. Breaker, “Widespread Genetic Switches and Toxicity Resistance Proteins for Fluoride,” *Science*, vol. 335, no. 6065, pp. 233–235, Jan. 2012, doi: 10.1126/science.1215063.
- [67] J. W. Nelson, R. M. Atilho, M. E. Sherlock, R. B. Stockbridge, and R. R. Breaker, “Metabolism of Free Guanidine in Bacteria is Regulated by a Widespread Riboswitch Class,” *Mol. Cell*, vol. 65, no. 2, pp. 220–230, Jan. 2017, doi: 10.1016/j.molcel.2016.11.019.
- [68] M. E. Sherlock, S. N. Malkowski, and R. R. Breaker, “Biochemical Validation of a Second Guanidine Riboswitch Class in Bacteria,” *Biochemistry*, vol. 56, no. 2, pp. 352–358, Jan. 2017, doi: 10.1021/acs.biochem.6b01270.

- [69] M. E. Sherlock and R. R. Breaker, “Biochemical Validation of a Third Guanidine Riboswitch Class in Bacteria,” *Biochemistry*, vol. 56, no. 2, pp. 359–363, Jan. 2017, doi: 10.1021/acs.biochem.6b01271.
- [70] H. Salvail, A. Balaji, D. Yu, A. Roth, and R. R. Breaker, “Biochemical Validation of a Fourth Guanidine Riboswitch Class in Bacteria,” *Biochemistry*, Nov. 2020, doi: 10.1021/acs.biochem.0c00793.
- [71] F. Lenkeit, I. Eckert, J. S. Hartig, and Z. Weinberg, “Discovery and Characterization of a Fourth class of Guanidine Riboswitches,” *Nucleic Acids Res.*, vol. 48, no. 22, pp. 12889–12899, Dec. 2020, doi: 10.1093/nar/gkaa1102.
- [72] G. Mirihana Arachchilage, M. E. Sherlock, Z. Weinberg, and R. R. Breaker, “SAM-VI RNAs Selectively Bind S-adenosylmethionine and Exhibit Similarities to SAM-III Riboswitches,” *RNA Biol.*, vol. 15, no. 3, pp. 371–378, Mar. 2018, doi: 10.1080/15476286.2017.1399232.
- [73] S. F. Dowdy, “Overcoming Cellular Barriers for RNA Therapeutics,” *Nat. Biotechnol.*, vol. 35, no. 3, Art. no. 3, Mar. 2017, doi: 10.1038/nbt.3802.
- [74] J. Conde and N. Artzi, “Are RNAi and miRNA Therapeutics Truly Dead?,” *Trends Biotechnol.*, vol. 33, no. 3, pp. 141–144, Mar. 2015, doi: 10.1016/j.tibtech.2014.12.005.
- [75] X. Hou, T. Zaks, R. Langer, and Y. Dong, “Lipid Nanoparticles for mRNA Delivery,” *Nat. Rev. Mater.*, vol. 6, no. 12, Art. no. 12, Dec. 2021, doi: 10.1038/s41578-021-00358-0.

- [76] E. Loh, F. Righetti, H. Eichner, C. Twittenhoff, and F. Narberhaus, “RNA Thermometers in Bacterial Pathogens,” *Microbiol. Spectr.*, Apr. 2018, doi: 10.1128/microbiolspec.RWR-0012-2017.
- [77] S. Dvir *et al.*, “Deciphering the Rules by which 5'-UTR Sequences Affect Protein Expression in Yeast,” *Proc. Natl. Acad. Sci.*, vol. 110, no. 30, Jul. 2013, doi: 10.1073/pnas.1222534110.
- [78] J. c. Nshogozabahizi, K. l. Aubrey, J. a. Ross, and N. Thakor, “Applications and Limitations of Regulatory RNA Elements in Synthetic Biology and Biotechnology,” *J. Appl. Microbiol.*, vol. 127, no. 4, pp. 968–984, 2019, doi: 10.1111/jam.14270.
- [79] V. Panchal and R. Brenk, “Riboswitches as Drug Targets for Antibiotics,” *Antibiotics*, vol. 10, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/antibiotics10010045.
- [80] I. Artsimovitch and T. M. Henkin, “*In vitro* Approaches to Analysis of Transcription Termination,” *Methods*, vol. 47, no. 1, pp. 37–43, Jan. 2009, doi: 10.1016/j.ymeth.2008.10.006.
- [81] R. O. Niederer, M. F. Rojas-Duran, B. Zinshteyn, and W. V. Gilbert, “Direct Analysis of Ribosome Targeting Illuminates Thousand-fold Regulation of Translation Initiation,” *Cell Syst.*, Jan. 2022, doi: 10.1016/j.cels.2021.12.002.
- [82] M. A. Griffin, J. H. Davis, and S. A. Strobel, “Bacterial Toxin RelE: A Highly Efficient Ribonuclease with Exquisite Substrate Specificity Using Atypical Catalytic Residues,” *Biochemistry*, vol. 52, no. 48, pp. 8633–8642, Dec. 2013, doi: 10.1021/bi401325c.

- [83] C. W. Reiss and S. A. Strobel, “Structural Basis for Ligand Binding to the Guanidine-II Riboswitch,” *RNA*, vol. 23, no. 9, pp. 1338–1343, Sep. 2017, doi: 10.1261/rna.061804.117.
- [84] L. Huang, J. Wang, and D. M. J. Lilley, “The Structure of the Guanidine-II Riboswitch,” *Cell Chem. Biol.*, doi: 10.1016/j.chembiol.2017.05.014.
- [85] A. Serganov *et al.*, “Structural Basis for Discriminative Regulation of Gene Expression by Adenine- and Guanine-Sensing mRNAs,” *Chem. Biol.*, vol. 11, no. 12, pp. 1729–1741, Dec. 2004, doi: 10.1016/j.chembiol.2004.11.018.
- [86] M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, and R. R. Breaker, “Riboswitches Control Fundamental Biochemical Pathways in *Bacillus subtilis* and Other Bacteria,” *Cell*, vol. 113, no. 5, pp. 577–586, May 2003, doi: 10.1016/S0092-8674(03)00391-X.
- [87] S. D. Gilbert, C. D. Stoddard, S. J. Wise, and R. T. Batey, “Thermodynamic and Kinetic Characterization of Ligand Binding to the Purine Riboswitch Aptamer Domain,” *J. Mol. Biol.*, vol. 359, no. 3, pp. 754–768, Jun. 2006, doi: 10.1016/j.jmb.2006.04.003.
- [88] M. Mandal and R. R. Breaker, “Adenine Riboswitches and Gene Activation by Disruption of a Transcription Terminator,” *Nat. Struct. Mol. Biol.*, vol. 11, no. 1, pp. 29–35, Jan. 2004, doi: 10.1038/nsmb710.
- [89] S. Pestka, “Studies on the Formation of Transfer Ribonucleic Acid-Ribosome Complexes,” *J. Biol. Chem.*, vol. 243, no. 15, pp. 4038–4044, Aug. 1968, doi: 10.1016/S0021-9258(18)93276-X.

- [90] S. Guedich *et al.*, “Quantitative and Predictive Model of Kinetic Regulation by *E. coli* TPP Riboswitches,” *RNA Biol.*, vol. 13, no. 4, pp. 373–390, Apr. 2016, doi: 10.1080/15476286.2016.1142040.
- [91] A. D. Garst, A. L. Edwards, and R. T. Batey, “Riboswitches: Structures and Mechanisms,” *Cold Spring Harb. Perspect. Biol.*, vol. 3, no. 6, Jun. 2011, doi: 10.1101/cshperspect.a003533.
- [92] A. Peselis and A. Serganov, “Themes and Variations in Riboswitch Structure and Function,” *Biochim. Biophys. Acta BBA - Gene Regul. Mech.*, vol. 1839, no. 10, pp. 908–918, Oct. 2014, doi: 10.1016/j.bbagr.2014.02.012.
- [93] K. E. Watters, E. J. Strobel, A. M. Yu, J. T. Lis, and J. B. Lucks, “Cotranscriptional Folding of a Riboswitch at Nucleotide Resolution,” *Nat. Struct. Mol. Biol.*, vol. advance online publication, Oct. 2016, doi: 10.1038/nsmb.3316.
- [94] R. Rieder, K. Lang, D. Graber, and R. Micura, “Ligand-Induced Folding of the Adenosine Deaminase A-Riboswitch and Implications on Riboswitch Translational Control,” *ChemBioChem*, vol. 8, no. 8, pp. 896–902, May 2007, doi: 10.1002/cbic.200700057.
- [95] V. de Jesus, J. Schmid, and B. Fürtig, “Binding of 30S Ribosome Induces Single-stranded Conformation Within and Downstream of the Expression Platform in a Translational Riboswitch,” *J. Mol. Biol.*, p. 167668, Jun. 2022, doi: 10.1016/j.jmb.2022.167668.
- [96] E. J. Strobel, L. Cheng, K. E. Berman, P. D. Carlson, and J. B. Lucks, “A Ligand-Gated Strand Displacement Mechanism for ZTP Riboswitch Transcription Control,”

- Nat. Chem. Biol.*, vol. 15, no. 11, Art. no. 11, Nov. 2019, doi: 10.1038/s41589-019-0382-7.
- [97] T. Schamber, O. Binas, A. Schlundt, A. Wacker, and H. Schwalbe, “Characterization of Structure and Dynamics of the Guanidine-II Riboswitch from *Escherichia coli* by NMR Spectroscopy and Small-Angle X-ray Scattering (SAXS),” *ChemBioChem*, vol. 23, no. 3, p. e202100564, 2022, doi: 10.1002/cbic.202100564.
- [98] D. Funck *et al.*, “Discovery of a Ni<sup>2+</sup>-Dependent Guanidine Hydrolase in Bacteria,” *Nature*, vol. 603, no. 7901, Art. no. 7901, Mar. 2022, doi: 10.1038/s41586-022-04490-x.
- [99] H. H. Guo, J. Choe, and L. A. Loeb, “Protein Tolerance to Random Amino Acid Change,” *Proc. Natl. Acad. Sci.*, vol. 101, no. 25, pp. 9205–9210, Jun. 2004, doi: 10.1073/pnas.0403255101.
- [100] M. Kwon and S. A. Strobel, “Chemical Basis of Glycine Riboswitch Cooperativity,” *RNA*, vol. 14, no. 1, pp. 25–34, Jan. 2008, doi: 10.1261/rna.771608.
- [101] J. J. Trausch, P. Ceres, F. E. Reyes, and R. T. Batey, “The Structure of a Tetrahydrofolate-Sensing Riboswitch Reveals Two Ligand Binding Sites in a Single Aptamer,” *Structure*, vol. 19, no. 10, pp. 1413–1423, Dec. 2011, doi: 10.1016/j.str.2011.06.019.
- [102] A. Gao and A. Serganov, “Structural Insights into Recognition of c-di-AMP by the *ydaO* Riboswitch,” *Nat. Chem. Biol.*, vol. 10, no. 9, pp. 787–792, Sep. 2014, doi: 10.1038/nchembio.1607.

- [103] M. I. Stefan and N. Le Novère, “Cooperative Binding,” *PLoS Comput. Biol.*, vol. 9, no. 6, p. e1003106, Jun. 2013, doi: 10.1371/journal.pcbi.1003106.
- [104] Z. Weinberg, J. W. Nelson, C. E. Lünse, M. E. Sherlock, and R. R. Breaker, “Bioinformatic Analysis of Riboswitch Structures Uncovers Variant Classes with Altered Ligand Specificity,” *Proc. Natl. Acad. Sci.*, vol. 114, no. 11, pp. E2077–E2085, Mar. 2017, doi: 10.1073/pnas.1619581114.
- [105] M. E. Sherlock, N. Sudarsan, and R. R. Breaker, “Riboswitches for the Alarmone ppGpp Expand the Collection of RNA-Based Signaling Systems,” *Proc. Natl. Acad. Sci.*, vol. 115, no. 23, pp. 6052–6057, Jun. 2018, doi: 10.1073/pnas.1720406115.
- [106] M. E. Sherlock, N. Sudarsan, S. Stav, and R. R. Breaker, “Tandem Riboswitches Form a Natural Boolean Logic Gate to Control Purine Metabolism in Bacteria,” *eLife*, vol. 7, p. e33908, Mar. 2018, doi: 10.7554/eLife.33908.
- [107] M. E. Sherlock, H. Sadeeshkumar, and R. R. Breaker, “Variant Bacterial Riboswitches Associated with Nucleotide Hydrolase Genes Sense Nucleoside Diphosphates,” *Biochemistry*, vol. 58, no. 5, pp. 401–410, Feb. 2019, doi: 10.1021/acs.biochem.8b00617.
- [108] R. M. Atilho, K. R. Perkins, and R. R. Breaker, “Rare Variants of the FMN Riboswitch class in *Clostridium difficile* and Other Bacteria Exhibit Altered Ligand Specificity,” *RNA*, p. rna.067975.118, Oct. 2018, doi: 10.1261/rna.067975.118.
- [109] S. Hamal Dhakal, S. S. S. Panchapakesan, P. Slattery, A. Roth, and R. R. Breaker, “Variants of the Guanine Riboswitch Class Exhibit Altered Ligand Specificities for Xanthine, Guanine, or 2'-deoxyguanosine,” *Proc. Natl. Acad. Sci.*, vol. 119, no. 22, p. e2120246119, May 2022, doi: 10.1073/pnas.2120246119.

- [110] C. A. Kellenberger *et al.*, “GEMM-I riboswitches from *Geobacter* Sense the Bacterial Second Messenger Cyclic AMP-GMP,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 112, no. 17, pp. 5383–5388, Apr. 2015, doi: 10.1073/pnas.1419328112.
- [111] D. E. Andreev, J. Hirnet, I. M. Terenin, S. E. Dmitriev, M. Niepmann, and I. N. Shatsky, “Glycyl-tRNA Synthetase Specifically Binds to the Poliovirus IRES to Activate Translation Initiation,” *Nucleic Acids Res.*, vol. 40, no. 12, pp. 5602–5614, Jul. 2012, doi: 10.1093/nar/gks182.
- [112] J. L. Brunelle, E. M. Youngman, D. Sharma, and R. Green, “The Interaction between C75 of tRNA and the A Loop of the Ribosome Stimulates Peptidyl Transferase Activity,” *RNA*, vol. 12, no. 1, pp. 33–39, Jan. 2006, doi: 10.1261/rna.2256706.
- [113] S. E. Walker and K. Fredrick, “Preparation and Evaluation of Acylated tRNAs,” *Methods*, vol. 44, no. 2, pp. 81–86, Feb. 2008, doi: 10.1016/j.ymeth.2007.09.003.
- [114] M. Martin, “Cutadapt Removes Adapter Sequences from High-Throughput Sequencing Reads,” *EMBnet.journal*, vol. 17, no. 1, pp. 10–12, May 2011, doi: 10.14806/ej.17.1.200.
- [115] B. Langmead and S. L. Salzberg, “Fast Gapped-Read Alignment with Bowtie 2,” *Nat. Methods*, vol. 9, no. 4, pp. 357–359, Apr. 2012, doi: 10.1038/nmeth.1923.
- [116] A. R. Gruber, R. Lorenz, S. H. Bernhart, R. Neuböck, and I. L. Hofacker, “The Vienna RNA Websuite,” *Nucleic Acids Res.*, vol. 36, no. suppl\_2, pp. W70–W74, Jul. 2008, doi: 10.1093/nar/gkn188.