

Mechanistic insights into molecular recognition and the regulatory landscape of the lysine riboswitch

by

By Andrew Garst

B.S. Biology
University of New Mexico Highlands 2006

Submitted to the faculty of the Graduate
School of the University of Colorado in partial
fulfillment of the requirement for the degree of
Doctor of Philosophy
Department of Chemistry and Biochemistry
2012

This thesis entitled:

**Mechanistic insights into molecular recognition and the regulatory landscape of
the lysine riboswitch
written by Andrew D. Garst
has been approved for the Department of Chemistry and Biochemistry**

Robert Batey

Deborah Wuttke

Amy Palmer

Rob Knight

Michael Yarus

Date _____

**The final copy of this thesis has been examined by the signatories, and we find
that both the content and the form meet acceptable presentation standards of
scholarly work in the above mentioned discipline.**

Garst, Andrew D. (Ph.D., Chemistry and Biochemistry)
Mechanistic insights into molecular recognition and the regulatory landscape of the
lysine riboswitch
Thesis directed by Associate Professor Robert T. Batey

RNA plays a central role in gene regulation and information processing in all kingdoms of life. Found in 5' leader sequence of many bacterial mRNAs, riboswitches serve as an exemplar of this duality as they control expression of their own transcript by directly binding small molecule metabolites in the cell. These RNAs adopt tertiary structures to scaffold highly specific ligand binding sites, reminiscent of their protein counterparts. Ligand binding is then coupled to conformational changes in the RNA that influence expression of the downstream message by altering the transcription or translation of the message.

To investigate the structural basis by which bacterial mRNAs specifically recognize lysine the aptamer domain from the leader sequence of the *asd* gene in *Thermotoga maritima* was solved by X-ray crystallography in both the free and bound conformations. These structures were complemented with solution based approaches to demonstrate that the tertiary architecture of the lysine aptamer domain is largely preorganized at 5 mM Mg^{2+} in the absence of ligand. Ligand binding was found to induce limited conformational changes within the five-way junction of the RNA. Based on these collective observations a site-specifically labeled RNA construct was designed to enable further thermodynamic and kinetic analysis of ligand binding to the aptamer. Using a series of lysine analogs that challenge key aspects of the structural model, we obtained a detailed understanding of the energetics of ligand recognition and demonstrate the importance of solvent and ion-mediated contacts in achieving a high affinity interaction. The binding kinetics of these analogs were also used to develop a simple mathematical framework for predicting the regulatory behavior of the RNA during transcription.

Kinetic predictions were tested using a minimally reconstituted *in vitro* transcription assay to gain further empirical insight into the regulatory functions of the *B. subtilis* lysine riboswitch and correlate the biological function with studies of the isolated aptamer. The regulatory response of lysine was found to agree well with a simple two state mechanism of ligand binding for lysine at a variety of NTP concentrations. The five fold variation in T_{50} observed for lysine along with changes in the observed termination efficiency also suggest that this RNA by indirect means integrates a more global picture of metabolism into its regulatory response. Kinetic predictions were also predictive for the regulatory response of many of the alternative ligands at low NTP concentrations, but were found to be less accurate in predicting responses at elevated NTP concentrations, suggesting that the simple model may neglect certain features of the transcription process.

The *in vitro* transcription assays were also employed to study the mechanism by which ligand binding is coupled to the secondary structural switch in the expression platform. A systematic survey of mutations to the P1 helix demonstrated that this element serves as the primary module for interdomain communication in the natural riboswitches, an insight that facilitated approaches to rationally design and optimize chimeric riboswitches. These studies have collectively shown that the regulatory switch

is self contained in the expression platform and can be reprogrammed to be responsive to a large number of alternative ligands through a simple mix and match approach.

Acknowledgements

I am grateful to many people for their contributions and support in this thesis. I would like to start by thanking my advisor Rob Batey to whom I am indebted for having maintained an approachable mentorship style and contagious enthusiasm for science. Rob provided a highly motivating environment to explore my scientific interests in an exciting field of research. I can readily attribute much of my graduate success to his many insights and suggestions.

I would also like to thank my committee members Deborah Wuttke, Amy Palmer, Rob Knight and Michael Yarus for their critical evaluations of my research and guidance during the course of my graduate career. Their kind support and letters of recommendation have been of great assistance in furthering my future scientific interests. Thank you.

I would not be writing this had I not had so many inspiring and supportive teachers throughout my *long* and... illustrious (?) educational career. Of the people I had the pleasure of knowing at New Mexico Highlands University I would especially like to thank Ben Nelson for having introduced me to scientific research through a project in his lab. His encouragement both in and out of the classroom played an important part in my decision to pursue a graduate education. I would like to thank Carol Linder for her excitement in the classroom, and for her dependable support and advice. I would also like to thank Eddie Tafoya, Merritt Helvenston and Dick Greene for their advice and intellectually enriching conversations and comedic style. As a graduate student I have had the sincere pleasure of interacting with many passionate scientists whose earnest dedication to the dissemination of knowledge, both in and out of the classroom, is an invaluable asset of the Boulder campus. I would like to specifically thank Art Pardi, Amy Palmer, Rob Knight, and Norm Pace for their clear and helpful classroom presentations, and the evident love they all share for teaching.

I have been fortunate to have befriended a number of extraordinary people in my time here at CU Boulder, people insightful and interesting perspectives on both life and science have enriched my own thinking. This list includes, but is not limited to (and not necessarily in order of importance) Colby Stoddard, Andrea Edwards, Lisa Warner, Kyle Landgraf, Ely Porter, Mirvat Abdelhaq, Pablo Ceres and James J. Johnson. These people in particular were central to my experience as a graduate student. I would also like to acknowledge Michael Vigil and Parker Micek, who's enduring friendship has been a defining aspect of my personal life.

Saving the best for last, I would like to dedicate this thesis to the various members of my loving family. To my wife Amber Garst whom has been a strong and supportive partner throughout the better part of the last decade, and made innumerable sacrifices on my behalf. Without her, none of this would be possible. To my two amazing sons Athan and Alakai with whom I look forward to sharing many deep philosophical conversations as they reach their teenage years. To my siblings Erin, Mike and Ben, whom have all graciously endured their older brother. To my mother Karen Reedy, whose encouragement and support I have to thank for my love of learning. And finally to my father, Jon Garst who taught me the value of work hard, and who taught me the importance of embracing life through the good times and the bad.

Table of Contents

Abstract	iii
Acknowledgements	v
Table of Figures	viii
Chapter 1: Introduction	1
1.1- Diverse regulatory roles of RNA in biology	1
1.2- Riboswitch distribution and regulation	4
1.3- Versatile mechanisms of riboswitch regulation	6
1.4- Structural themes of riboswitch aptamer domains	10
1.5- The riboswitch folding landscape: a case study of the purine riboswitch	14
1.4- Determinants of riboswitch specificity	18
1.5 Conformational changes associated with ligand binding	22
1.6 Temporal considerations for transcriptional regulation	24
1.7 Synthetic RNAs as a model for secondary structural switching	26
1.8- Regulation of lysine metabolism by riboswitches	29
Chapter 2: Structural basis for recognition of lysine by mRNA.	36
2.1- Crystallization and structural determination of the asd aptamer from <i>Thermatoga maritima</i>	36
2.2- Global Organization of the lysine riboswitch	41
2.3- Lysine recognition	44
2.4- Ligand-dependent conformational changes in the RNA	47
2.5- Temperature dependent folding of the lysine riboswitch	54
2.6- Resistance mutations in the lysine riboswitch	57
2.7- Materials and Methods	60
Chapter 3: Correlative examination of the thermodynamics, kinetics and regulatory response of the lysine riboswitch	63
3.1- Characterization and validation of a 2-aminopurine labeled receptor.	65
3.2- Molecular determinants of specificity	69
3.3- Evidence for rigidity of the lysine binding site	74
3.4- Binding kinetics suggest a rapid approach to equilibrium.	79
3.5- Assessing the effects of transcription speed on regulation.	85

3.5- <i>Discussion</i>	90
3.6- <i>Methods and Materials</i>	94
Chapter 4: Rational design and optimization of chimeric riboswitches	97
4.1- <i>Design strategy for generating chimeric riboswitches</i>	101
4.2- <i>Wild type riboswitches as a benchmark for comparative analysis</i>	101
4.3- <i>Functional characteristics of chimeric riboswitches</i>	104
4.4- <i>Mutational analysis of the P1 element</i>	108
4.4- <i>Chimeric riboswitches are functional in vivo.</i>	113
4.5- <i>Discussion.</i>	115
4.6- <i>Methods and Materials</i>	118
Chapter 5: Conclusions and future directions	120
5.1 <i>Structural similarities between lysine recognition by RNA and proteins</i>	120
5.2 <i>Riboswitches in the cellular milieu: where do we go from here?</i>	124
Works Cited	128

Table of Figures

Figure 1.1	Distribution of common riboswitch classes in the bacterial kingdom.....	5
Figure 1.2	Logic and mechanism of riboswitch regulation.....	7
Figure 1.3	Diverse modes of riboswitch regulation.....	9
Figure 1.4	Recurrent structural themes of functional RNAs.....	12
Figure 1.5	Folding of purine riboswitch aptamer domain.....	17
Figure 1.6	Examples of riboswitch specificity for purines.....	20
Figure 1.7	Comparison of solvent inaccessible surface area for ligands bound to naturally or artificially derived aptamers.....	21
Figure 1.8	Models for cotranscriptional folding of synthetic and natural RNA switches.....	28
Figure 1.9	Diaminopimelate pathway for lysine biosynthesis.....	32
Figure 1.10	Consensus structure and specificity profile of the lysine riboswitch.....	33
Figure 2.1	Crystal construct design strategy for the <i>T. maritima</i> asd aptamer.....	37
Figure 2.2	Structural determination of the lysine riboswitch.....	39
Figure 2.3	Global view of lysine riboswitch tertiary structure.....	42
Figure 2.4	Structural motifs that stabilize global architecture of the lysine riboswitch.....	43
Figure 2.5	Details of lysine recognition.....	45
Figure 2.6	Comparison of bound and free state structures observed by X-ray crystallography.....	48
Figure 2.7	Conformational changes associated with Mg^{2+} and lysine binding.....	50
Figure 2.8	Sequence and structure of the RNAs used in SHAPE probing experiments.....	51
Figure 2.9	SAXS scattering curves of the lysine riboswitch.....	53
Figure 2.10	Temperature dependence of the lysine riboswitch folding.....	56
Figure 2.11	Mutations that confer AEC resistance.....	59
Figure 3.1	Validation of the 2AP reporter for determination of ligand binding affinities.....	67
Figure 3.2	Important contacts of the lysine riboswitch binding pocket.....	68
Figure 3.3	Test of standard amino acids for binding competency.....	71
Figure 3.4	Energetics of molecular recognition.....	72
Figure 3.5	Chemical probing of the lysine riboswitch.....	75
Figure 3.6	Lysine binding pocket enforces strict maintenance of H-bonding interactions.....	78
Figure 3.7	Analysis of lysine binding kinetics by stopped flow fluorescence spectroscopy.....	81
Figure 3.8	Sequence and structure of the <i>lysC</i> riboswitch.....	82
Figure 3.9	Effects of transcriptional speed and ligand binding on transcriptional regulation.....	87
Figure 3.10	K_D and T_{50} comparisons for lysine and lysine analogs.....	89
Figure 4.1	Apparent modularity of naturally occurring expression platforms.....	100
Figure 4.2	Outline of chimeric riboswitch design strategy.....	102
Figure 4.3	Response characteristics of reprogrammed riboswitches.....	106
Figure 4.4	Analysis of P1 stability on the regulatory response of the <i>lysC</i> riboswitch.....	110
Figure 4.5	Modulation of P1 provides a general tuning mechanism.....	112
Figure 4.6	In vivo regulatory responses of chimeric RNAs.....	114
Figure 5.1	Comparison of lysine binding pockets of the lysine riboswitch and <i>lysyl-tRNA synthetase</i>.....	122

Chapter 1: Introduction

1.1- Diverse regulatory roles of RNA in biology

The survival of biological organisms is predicated on their ability to rapidly adapt to physical and chemical changes in the environment. To orchestrate this process cells employ a sophisticated network of regulatory mechanisms that modulate gene expression and tune the metabolic state of the cell. It has long been appreciated that regulatory proteins such as transmembrane receptors and transcription factors are involved in coordinating adaptive physiological responses based on extracellular cues. RNA on the other hand has only recently begun to receive recognition for its diverse regulatory roles in biology, despite longstanding knowledge of its central role in the process of information transfer, particularly in the decoding of genetic information during translation^{1,2}.

Recognition of RNA regulation has been spurred in large part by the discovery of short (22 nucleotide) RNA known as micro RNAs (miRNAs) and silencing RNAs (siRNAs) in eukaryotes that influence the translation or degradation of their complementary mRNA targets³⁻⁵. Endogenously expressed miRNAs for example have found to play fundamental roles coordinating the developmental programs of many metazoan organisms, and are estimated to control the expression of up to ~ 50% of all mammalian mRNAs⁶. More recently many novel classes of long non-coding RNAs (>200 nucleotides) in eukaryotes have been found to play important functional roles in directing chromatin modification, regulation of transcription initiation, and the control of mRNA processing, editing and degradation⁷. This functional diversity may indicate that a

vast network of regulatory RNA functions remain to be elucidated⁸. Indeed it has been suggested that increased morphological complexity of metazoans may be accounted for by the corresponding increase in non-coding potential of in these organisms⁹.

While the proportion of non-coding sequence in bacterial genomes is much smaller than that of their eukaryotic counterparts⁹, they have evolved equally stunning forms of RNA based regulation to cope with a wide variety of environmental stresses. For example bacteria express a host of endogenous short RNAs (sRNAs) that alter translation or stability of their target messages in a fashion analogous (but not homologous) to eukaryotic miRNAs. These sRNAs coordinate adaptive responses to many stimuli including Fe^{2+} limitation, carbon shortage, oxidative stress, low temperatures and in quorum sensing^{10,11}. To cope with the abundance of bacteriophage in their natural environments, bacteria (and archaea) have also evolved RNA based forms of immunity that makes use of short RNAs to target the destruction of foreign phage derived nucleic acids¹². These transcripts are derived from Clustered Regularly Interspaced Short Palindromic Repeat or CRISPR loci that are targeted for cleavage by a series of CRISPR associated proteins (CASP) to generate a mature RNA species that can specifically base pair with exogenous nucleic acid sequences¹². This elegant system not only targets RNA and DNA species for destruction but also serves as the basis for a simple form of molecular memory by incorporating exogenous sequences into the CRISPR loci to effectively immunize the bacteria against related phage, providing the first evidence that these organisms have adaptive immune capabilities^{12,13}.

The specificity of many RNA based regulatory processes is conferred primarily by the base pairing between the regulatory RNA and its complementary message, as

indicated for the CRISPR RNAs above. Because base pairing follows a simple and reproducible pattern, many target messages can be readily identified from database searches that scan for complementary sequence targets. This feature of RNA regulation has provided a powerful tool for biologists to manipulate expression and cellular phenotypes with unprecedented accuracy, a fact that underlies the popularity of using siRNAs in functional screening efforts¹⁴. From the standpoint of the RNA however, these RNA-RNA based regulatory mechanisms do not necessarily take full advantage of the functional capacity of this remarkable polymer. For example *in vitro* selection studies starting in the 1980's had demonstrated that functional RNA sequences can be readily enriched from large random sequence libraries to perform the functions of small molecule or protein binding (performed by so-called RNA aptamers), or even catalysis (ribozymes)¹⁵. These results rely in part on the ability of RNA to adopt sophisticated tertiary structures, and in the last few decades a vast array of naturally occurring RNA elements have been demonstrated to take advantage of this property to perform naturally evolved biological functions¹. While examples are abound, some of the most remarkable exhibitions of RNA's functional capabilities are found in *cis*-acting elements in the leader sequence of bacterial mRNAs known as riboswitches¹⁶. These RNAs have naturally evolved to directly sense their chemical environment and relay this information to control gene expression. Detailed understanding of the mechanism by which natural regulatory RNAs couple ligand binding to regulation holds promise for guiding efforts to design increasingly more effective methods for incorporating *in vitro* derived aptamers into the creation of ligand responsive RNA based regulatory devices. The development of such techniques should not only continue to expand use of RNA in controlling and

monitoring complex biological systems, but enable deeper understanding of the principles that dictate the biological function of these versatile regulatory molecules.

1.2- Riboswitch distribution and regulation

The regulatory capacity of RNA is readily illustrated by cis-acting elements found in 5'-UTRs of bacterial mRNAs that respond to physiological stimuli without the aid of proteins, thereby allowing mRNA to direct its own expression¹⁶. There are a large variety of such elements that allow mRNAs to directly respond to many discrete environmental cues including changes in temperature, levels of uncharged tRNA, and the concentration of specific metal ions or small molecule metabolites¹⁷. The primary commonality of all of these regulatory mechanisms, including those that protein-RNA interactions, is the presence of a sequence that adopts one of two mutually exclusive secondary structures that lead to expression or repression of the parent transcript¹⁸. This folding decision occurs cotranscriptionally, with ligand binding steering the RNA into one of two folding pathways that ultimately dictates its fate.

Metabolite sensing RNAs, better known as riboswitches, represent a fundamental mechanism of gene regulation that is widespread in the bacterial kingdom (Figure 1.1), and is found controlling a significant proportion of gene expression in some organisms; ~4% genes in the *Bacillus subtilis*^{16,19}. In bacteria, they are always found in the 5' leader sequence although examples have been identified in the introns or 3' untranslated region (3'-UTR) in eukaryotic transcripts²⁰. There are currently at least twenty classes of riboswitches that recognize a broad range of ligands including purine nucleobases^{19,21}, amino acids²²⁻²⁵, vitamin cofactors²⁶⁻²⁹, aminosugars³⁰, metal ions^{31,32},

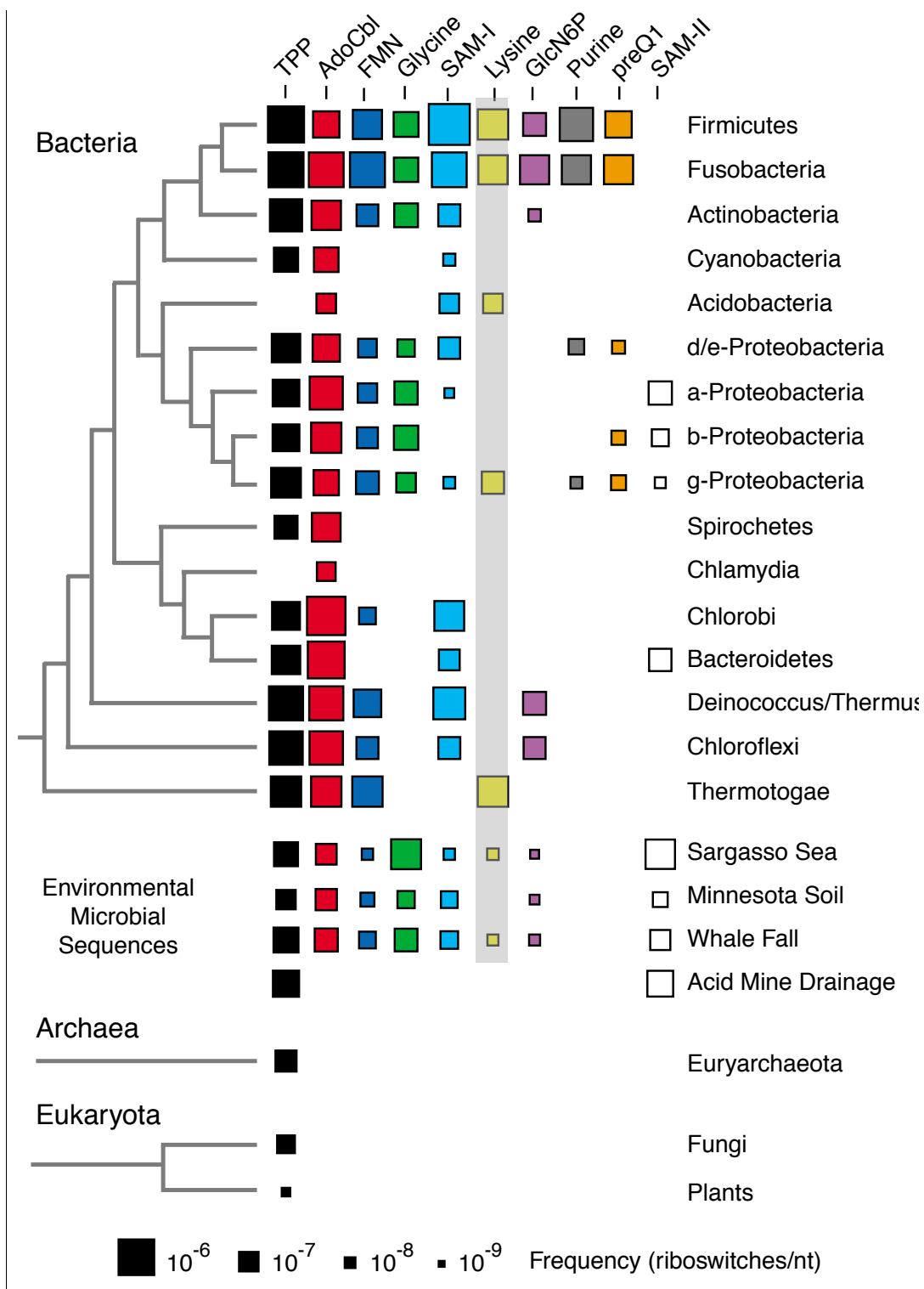


Figure 1.1 Distribution of common riboswitch classes in the bacterial kingdom.

The box size indicates the prevalence with which the riboswitch is found in public genomic databases. in the total genomic sequences available in The lysine class riboswitches are highlighted by the gray box as this class represents the focus of this work. Figure taken from ref³⁷. Abbreviations: TPP, thiamine pyrophosphate; AdoCbl, adenosylcobalamin; FMN, flavin mononucleotide; SAM, S-adenosylmethionine; GlcN6P glucosamine-6-phosphate.

and second messenger molecules³³. Many “orphan” classes have also been identified using phylogenetic analyses but their metabolites await identification³⁴⁻³⁶.

Riboswitches are typically composed of two distinct domains: a metabolite receptor known as the aptamer domain and an expression platform whose secondary structure signals the regulatory response (Figure 1.2). Due to the polarity of transcription, the aptamer domain (AD, Figure 1.2) is transcribed first, providing time for this receptor to sense the cellular environment before the expression platform (EP) is fully synthesized. Embedded within the aptamer domain is the switching sequence (Figure 1.2, sequence B), a sequence shared between the aptamer domain and expression platform. In the case of riboswitches that mediate transcription attenuation, ligand binding directs formation of a terminator hairpin (Figure 1.2, Term) causing the RNA polymerase to abort transcription prior to synthesizing the downstream coding sequence.

1.3- Versatile mechanisms of riboswitch regulation

While most riboswitches mediate transcription attenuation³⁷ their versatility is clearly demonstrated by their ability to regulate gene expression at different levels of the information transfer process (Figure 1.3). For example, many riboswitches couple ligand binding to the formation of secondary structures that sequester the ribosome binding site (RBS) within a helix, thereby allowing control of translation initiation³⁷, as was first demonstrated for the thiamine pyrophosphate (TPP) riboswitch²⁷(Figure 1.3A). Additionally, riboswitches can alter the rate of mRNA degradation by directing cleavage or splicing events that promote degradation of the RNA. The *glmS* riboswitch takes a

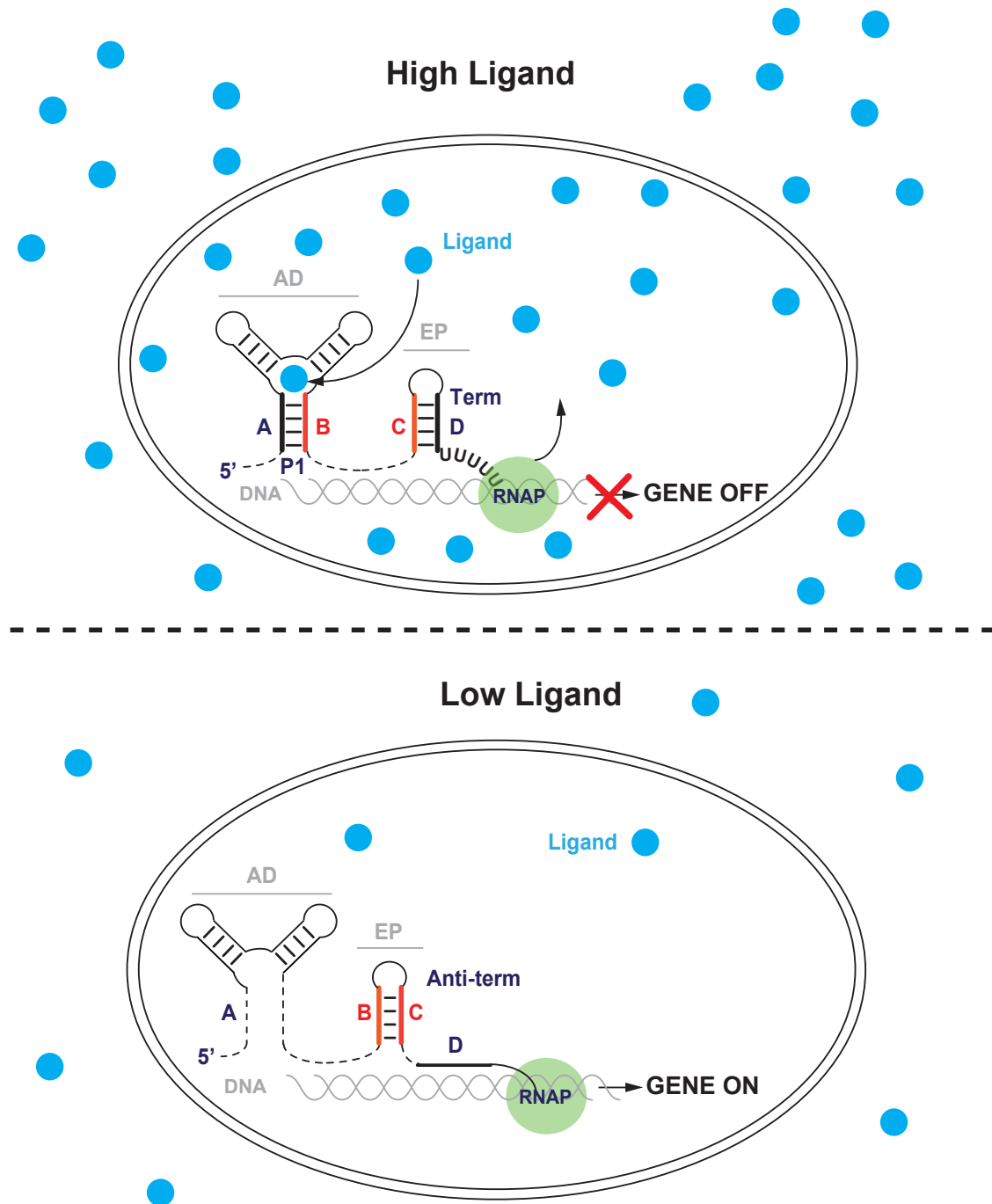


Figure 1.2 Logic and mechanism of riboswitch regulation.

Under high ligand concentrations the aptamer domain (AD, gray) productively binds its target metabolite stabilizing the P1 element formed between sequences A and B as indicated. Upon synthesis of the downstream expression platform (EP) a terminator hairpin (Term) is formed, directing the RNA polymerase (RNAP) to abort transcription of a gene that is typically involved in the biosynthesis or import of the metabolite. Under low ligand concentrations ligand binding does not occur and the EP directs unfolding of the P1 helix leading to formation of an antiterminator element. Gene expression is turned on under these circumstances allowing the metabolite to be produced by the cell.

unique approach to this by acting as a ribozyme that uses glucosamine-6-phosphate (GlcN6P) as an enzymatic cofactor that participates directly in cleavage of the 5' leader sequence resulting in destabilization of the transcript^{30,38,39} (Figure 1.3B). In contrast, eukaryotic thiamine pyrophosphate (TPP) riboswitches, found in the 3'-UTR or intergenic regions of specific thiamine biosynthesis genes, can control alternative splicing of the mRNA by sequestering splice sites within the riboswitch upon ligand binding, thereby sequestering these sites from the splicing machinery^{20,40} (Figure 1.3C). More recently, another mechanism of ligand dependent cleavage was discovered for a di-cyclic GMP (diGMP) riboswitch found upstream of a putative virulence gene in *Clostridium difficile*⁴¹. In this case, the aptamer was found to lie upstream of a Group I intron element that acts as the "expression platform." Binding of diGMP unmasks a cleavage site utilized by the intron to activate the self-cleavage activity of the intron, resulting in repression of the transcript⁴¹.

Although the majority of riboswitches provide relatively simple "ON" or "OFF" control points nature has found ways to arrange these RNAs in tandem to accomplish more sophisticated logic functions¹⁸. The simplest form of this involves a tandem arrangement consisting of two riboswitches that respond to the same ligand. The *gvcT* operon for example is controlled by a glycine riboswitch with two aptamer domains that cooperatively bind glycine, a remarkable feature that had only previously been ascribed to proteins^{18,30,42}. Cooperative binding allows a more digital response to small changes in ligand concentration that is unattainable by single copy riboswitches. In other cases the expression response can be coordinated by distinct metabolites, such as that found in the S-adenosylmethionine (SAM) and adenosylcobalamin (B12) responsive

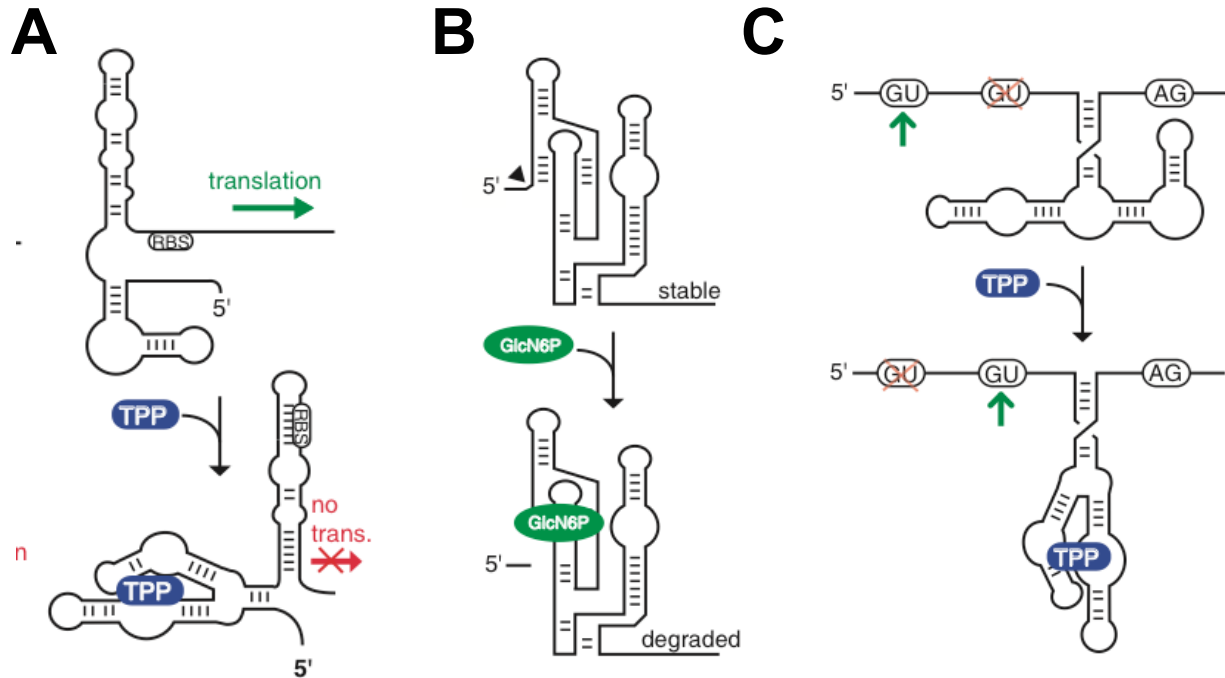


Figure 1.3 Diverse modes of riboswitch regulation

A) Model for TPP dependent regulation of translation initiation. Ligand binding sequesters the RBS thereby inhibiting translation initiation. B) The *glmS* riboswitch utilizes GlcN6P as a cofactor to mediate mRNA cleavage and subsequent destabilization. C) Model for alternative splicing mechanism mediated by the eukaryotic TPP riboswitch. In the absence of ligand, the RNA is spliced at the first splice site, creating an RNA that is inefficiently expressed due to the presence of upstream open reading frames that compete with the main ORF for translation. Ligand binding to the riboswitch induces conformational changes that expose a downstream 5' splice site, thereby upregulating expression of the ORF. Figures taken from ref¹⁹³.

elements⁴³. Each of these riboswitches can independently promote transcription termination upon ligand binding making this mRNA responsive to changes in the concentration of either metabolite. Riboswitches can also overlap with other post-transcriptional control mechanisms, as was recently described for the *Enterococcus faecalis eut* operon, whose expression allows ethanolamine utilization within the human gastrointestinal tract and is thought to affect the virulence of this organism⁴⁴. The operon's polycistronic transcript contains at least one B₁₂ riboswitch that terminates transcription in the presence of coenzyme B₁₂ and many copies of a conserved RNA

stem loop that is bound by a response regulatory protein to promote antitermination⁴⁴. In such cases, overlapping post-transcriptional regulatory mechanisms allow a signaling complexity rivaling that of DNA binding transcription factors that can finely tune gene expression from a specific promoter.

Interestingly, RNA switches have also recently been identified in the 3'-UTR of eukaryotic genes, leading to speculation that this region may provide a fertile ground for the discovery of novel mechanisms of riboregulation in these organisms. Indeed, a recently discovered RNA switch in the 3' untranslated region (UTR) of the human VEGFA mRNA forms mutually exclusive structures that alter the production of this growth factor during oxidative stress in myeloid cells⁴⁵. Although this regulatory element is not a true riboswitch, it does adopt mutually exclusive folds that interact with distinct protein partners to coordinate alternative expression responses. This process is driven by inhibiting proteasomal degradation of one protein during oxidative stress that allows for its intracellular accumulation, thus tipping the balance towards one of the two conformations. Mechanistic details of this process await further investigation, but this study highlights the continuing expansion of our knowledge of the scope of riboregulation in eukaryotic organisms.

1.4- Structural themes of riboswitch aptamer domains

Many antimetabolites have been shown to bind to riboswitch aptamers, suggesting that their toxicity can be at least in part attributed to their ability to repress riboswitch controlled pathways that are essential for bacterial survival or virulence⁴⁶. This observation has in part motivated efforts to obtain high-resolution models of riboswitch–

ligand complexes that explain the molecular details of ligand binding and provide clues into how these interactions are coupled to genetic regulation. These studies have revealed a number of themes common to the various riboswitches despite their obvious structural differences.

The most readily apparent feature of riboswitch aptamers has been the observation their seemingly diverse architectures are sculpted from the same set of recurrent RNA motifs found in initial structures of the ribosome and tRNA, thus confirming early predictions that the RNA “erector set” has been observed in its entirety⁴⁷. These motifs are responsible for introducing bending, distortion, and long-range packing between helices to enable the complicated tertiary architectures observed in functional RNAs^{48,49}, and include kink-turns⁵⁰, sarcin/ricin (or loopE) motifs⁵¹, GNRA tetraloops⁵², UUCG terminal loops⁵³ and A-minor motifs⁵⁴. One of the most common of these is the A-minor interaction in which an unpaired adenosine packs tightly against the minor groove of a distal helix to form a stable tertiary structure⁵⁴. There are nearly 100 A-minor motifs that participate in stabilizing the complicated architecture of the ribosome⁵⁴ and mediate critical aspects of group I intron folding^{55,56}. These motifs are abundant in riboswitches as well. is also conserved in particular subclasses of the di-cyclic GMP binding riboswitch. Similarly, A-minor motifs play important roles in the tertiary architecture of the GlcN6P, TPP, di-cyclic guanosine monophosphate (diGMP), and Mg²⁺ (Figure 1.4) riboswitches^{33,57-59}.

Other larger motifs can be co-opted to form the entire chassis of functional RNAs, including many riboswitch aptamers. For example many the architecture of many three-way junctions is stabilized by long-range loop-loop interactions including the purine

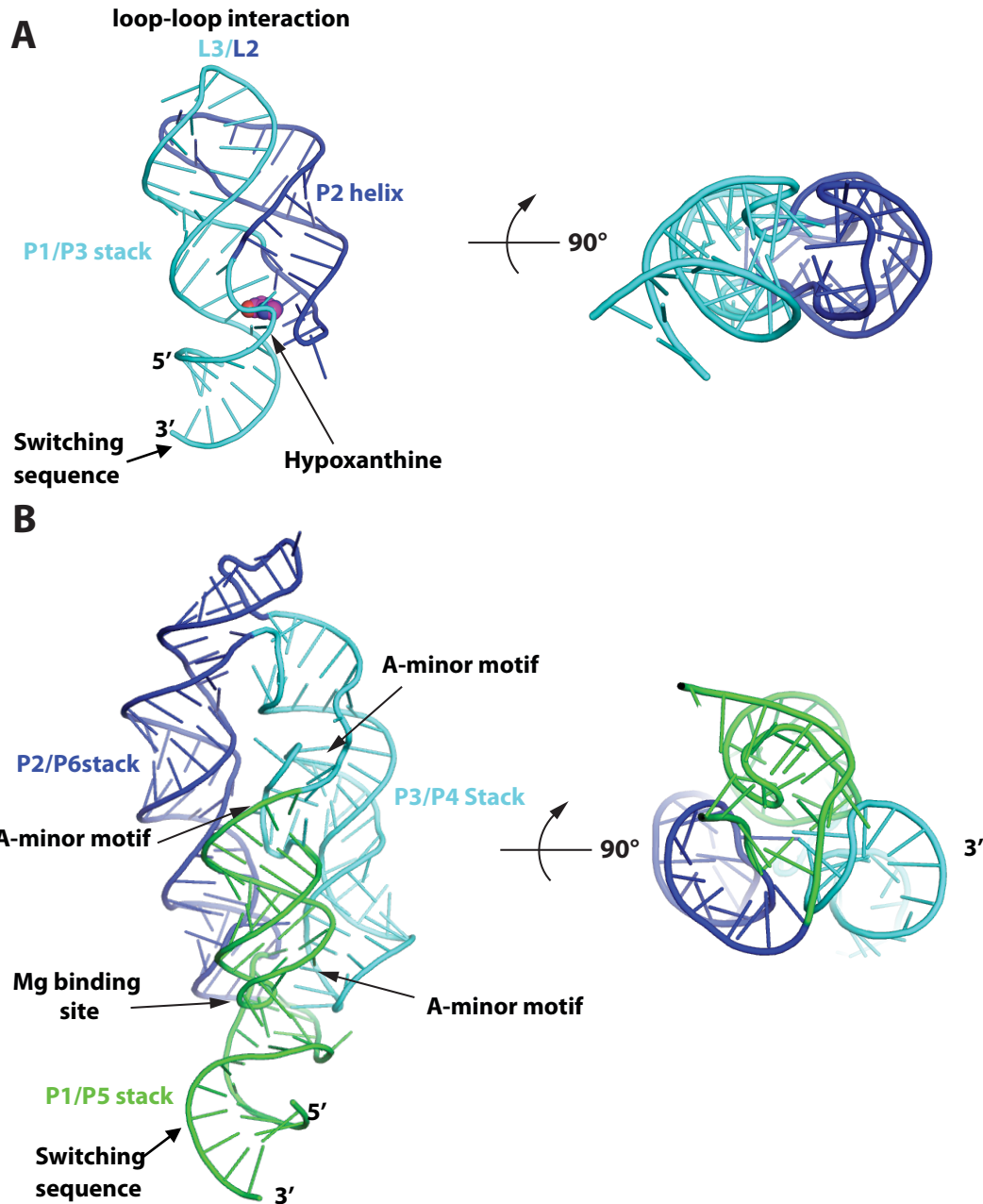


Figure 1.4 Recurrent structural themes of functional RNAs

The global structure of the (A) purine and (B) Mg^{2+} aptamer domains demonstrate recurrent structural themes of RNA. The purine riboswitch forms a two helix bundle (right) composed of the P1/P3 coaxial stack and the P2 helix and stabilized by a terminal loop-loop interaction between L2 and L3 segments. The Mg^{2+} riboswitch forms a three helix bundle stabilized by a number of A-minor motifs. Note the ligand binding sites (indicated by arrows) make contacts with the 3' side of the P1 helix which is typically involved in the secondary structural switch required for regulation.

riboswitch⁶⁰ (Figure 1.4 A), the hammerhead ribozyme⁶¹, and the thiamine pyrophosphate riboswitch^{62,63}, suggesting that this topology may be considered a fundamental unit of RNA structure^{64,65}. Likewise the classic H-type pseudoknot fold adopted by many viral RNAs that promote frameshifting during translation⁶⁶, and form a key functional element of telomerase RNA^{67,68} serves as the global scaffold of the SAM-II and preQ1 riboswitch aptamers⁶⁹⁻⁷¹. Surprisingly, riboswitch aptamers can even be constructed almost entirely from a single recurrent motif; the structure of an FMN binding riboswitch contains two symmetrical T-loop domains (P3–P5 and P2–P6) that align with the H19–H20 domain of the *Haloarcula marismortis* 23S rRNA with an r.m.s.d. of 1.48 Å, a striking similarity⁷².

In addition to the recurrence of defined structural motifs, riboswitches also share similar features in their global topology. For example, the helical elements of many aptamers are arranged to form coaxial stacks between contiguous helical elements (Figure 1.4). These stacks are further oriented into bundle like structures in which the coaxially stacked Watson–Crick paired helices are held parallel to one another by long-range tertiary interactions, typically mediated by the motifs mentioned above. In the SAM-I riboswitch for example, two coaxial helical stacks pack together to yield an arrangement similar to the catalytic core of group I introns^{24,73-75}. This structure is stabilized by a pseudoknot formed between the L2 loop and the J3/4 junction region⁷⁵. Like SAM-I, the purine and TPP riboswitches also form two helix bundles that are stabilized a terminal loop-loop and terminal loop-internal bulge interaction respectively (Figure 14A). The GlcN6P and Mg²⁺ binding aptamers on the other hand adopt more complicated three helix bundle arrangements.

Finally, almost all riboswitch structures have revealed that the switching sequence incorporated within the binding site or immediately adjacent, implying an intimate linkage between ligand binding and sequestration of this sequence within the aptamer domain (Figure 1.2, strand B). Indeed, in combination with chemical probing techniques, it is clear that ligand binding is accompanied by conformational changes around the binding pocket that often include the switching sequence. This concept was first uncovered through a structural analysis of the *B. subtilis xpt-pbuX* guanine riboswitch, a transcriptional regulator, in which ligand binding stabilizes formation of two base triples between residues in the three-way junction and the switching sequence⁶⁰. The SAM-II riboswitch has evolved an alternative solution in which the SAM ligand makes contacts with the ribosome binding sequence, thereby directly obstructing translation of the mRNA⁶⁹. This arrangement represents a rare case in which the aptamer domain and the expression platform are fully integrated into a single SAM binding pseudoknot. Thus, in every riboswitch that has been structurally characterized to date, ligand binding is directly coupled to conformational changes that involve the switching sequence, which forms the basis for their regulatory activity.

1.5- The riboswitch folding landscape: a case study of the purine riboswitch

Crystal structures of riboswitch aptamer domains are static pictures that do not necessarily lead to a direct understanding of the process by which these RNAs acquire their three-dimensional architecture. For riboswitches to function during transcription (Figure 1.2), they must rapidly form the tertiary interactions described earlier and avoid kinetic folding traps commonly found in RNA^{76,77}. In other words, slow folding or

misfolding will impede their ability to elicit a proper regulatory response during the short temporal window of transcription. While it is difficult to monitor the co-transcriptional folding process directly, a number of experimental techniques that induce denaturation by means of temperature, chemical, or force have enabled detailed analysis of the folding landscape of fully synthesized RNAs that allow some insight into the process by which these RNAs attain a native conformation.

Because of its simple architecture, the purine riboswitch has emerged as the dominant model system for understanding aptamer domain folding. A technique that uses the reactivity of 2'-hydroxyl groups to *N*-methylisatoic anhydride (NMIA) as a probe of RNA folding^{78,79} was recently used to study folding of the *B. subtilis xpt-pbuX* guanine riboswitch⁸⁰. An advantageous feature of this chemical probing method—commonly called “SHAPE”—is that it can be performed over a broad temperature range⁸¹, allowing for determination of the melting temperature (T_m , defined as the temperature at which the RNA is half folded and half unfolded) of every nucleotide in the RNA of interest⁸¹. Applied to the purine aptamer, at high temperatures ($\geq 70^\circ\text{C}$), only the helical elements of the aptamer are structured, as expected from the large free energy associated with RNA secondary structure formation (Figure 1.5A)⁸². As the temperature drops below 60°C , the L2-L3 interaction forms along with nucleotides becoming protected in J3/1, indicating that the peripheral loop-loop interaction promotes partial preorganization of the three-way junction via coaxial stacking of P1 and P3. Notably, J1/2 and J2/3 remain conformationally flexible in the absence of ligand, which would allow the effector access to the core of the RNA.

The folding analysis using SHAPE chemistry agrees well with single molecule fluorescence resonance energy transfer (smFRET) experiments with the *B. subtilis* *pbuE* adenine riboswitch⁸³. In this approach, folding transitions are reported by changes in the FRET value of single molecules in which L2 and L3 were labeled with donor and acceptor fluorophores, respectively (Figure 1.5A). Importantly, the smFRET data provide additional insight into the rates of folding and the influences of magnesium (Mg^{2+}) that is not possible using ensemble techniques. Addition of Mg^{2+} above 50 mM was shown to significantly increase the folding rate of the loop-loop interaction. At physiologically relevant Mg^{2+} concentrations, the global fold of the aptamer is adopted within ~ 1 s⁸³. Considering that bacterial polymerases have elongation rates of around ~ 45 nucleotides per second at $25^{\circ}C$ ⁸⁴, transcription of this 70 nucleotide aptamer would require about 1.5 s. Thus, folding and transcription occur on similar timescales, as would be expected for a cotranscriptional process.

Another method for monitoring folding processes uses force, which has gained popularity because of technological advances that allow a high degree of spatial and temporal resolution^{85,86}. Application of constant force to the ends of a macromolecule alters the energy landscape such that specific structural elements become bistable (equally occupying the folded and unfolded states) at a specific applied force, known as a $F_{1/2}$ ^{85,86}. Like smFRET, this technique simultaneously measures thermodynamic and kinetic parameters of folding and unfolding. Observations of the *Vibrio vulnificus* *add* adenine riboswitch folding by force spectroscopy revealed that as the applied force was decreased, extension changes corresponding to formation of P2 occur, followed by P3,

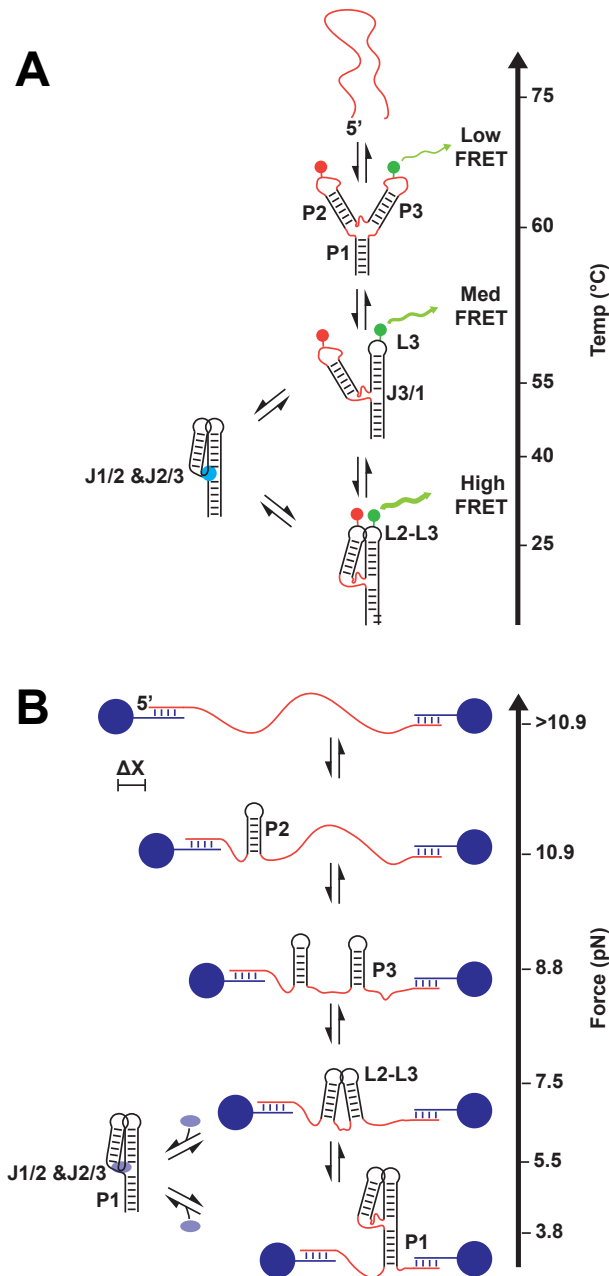


Figure 1.5 Folding of the purine riboswitch aptamer domain.

A) Thermal denaturation and smFRET of the purine riboswitch reveal similar hierarchical folding landscapes. The T_m of each structural element (denoted by red strands) of the xpt-pbuX purine riboswitch reveals folding intermediates similar to those observed in smFRET experiments of the pbuE riboswitch at ambient temperature. smFRET experiments consisted of a fluorescein donor (green) and Cy3 acceptor (red) placed at L2 and L3 respectively allow observation of the loop-loop interaction in real time. At ambient temperatures, NMR agrees well with the preformed L2-L3 conformation. B) Unfolding the pbuE riboswitch by force measured as changes in distance (DX). This experiment reveals a pathway in which individual helices form sequentially from 5' to 3'. Although denaturation studies show that ligand (cyan sphere) binding stabilizes the three way junction by 15°C, force spectroscopy supports a model in which P1 is also stabilized by ligand binding.

the loop-loop interaction, and finally P1 (Figure 1.5B)⁸⁷. By fully extending the RNA and allowing it to refold for predetermined time intervals before reapplying force, the authors obtained estimates for the rate of aptamer domain folding similar to those derived from smFRET measurements.

The folding model obtained by force spectroscopy is distinct from that obtained using other techniques⁸⁷. Although chemical probing and smFRET propose a classic hierarchical folding pathway in which all secondary structure is acquired before tertiary structure, the folding pathway proposed using force spectroscopy proceeds from the 5' to 3' direction. Such a model is more consistent with expectations for folding during transcription⁸⁸. Nevertheless, both models illustrate the importance of tertiary structure formation before ligand binding, as this has cooperative influences on ligand binding by pre-organizing the binding site.

1.4- Determinants of riboswitch specificity

The aptamer domain must accomplish two interconnected tasks: specific recognition of the effector and coupling binding to a regulatory outcome. The first objective requires that the RNA effectively discriminates between closely related compounds in the cell. For example, the purine riboswitches achieve over a 10,000-fold level of discrimination between guanine and adenine^{19,21}, whereas the lysine riboswitch discriminates between lysine and ornithine, amino acids that differ by a single methylene group in their side chain^{22,89}.

The purine riboswitch family has been the most well studied riboswitch to date and nicely demonstrates how a common RNA fold can achieve discrimination between

different purine nucleobases or nucleosides. The purine family can be divided into three sub-classes that are defined by their specific effector molecules: guanine/hypoxanthine, adenine, or 2'-deoxyguanosine⁹⁰. In each class, the effector is bound using a pocket formed by the RNA three-way junction^{19,21,91}. At the center is a base triple involving the ligand and two pyrimidine residues that serve to specifically recognize the nucleobase (Figure 1.6)^{60,92}. Every functional group in the ligand is directly recognized through hydrogen bonding interactions, in part explaining the ability of the RNA to achieve a high specificity for this compound. Surprisingly, the nucleobase does not completely stack on the bases above or below it, a feature common to other riboswitches that recognize nucleobase-containing effectors^{62,63,69,75,93}.

Purine specificity is primarily achieved through the identity of two pyrimidine residues that participate in a base triple with the ligand. The pyrimidine at position 74 discriminates between adenine and guanine or 2'-deoxyguanosine based on its ability to form a canonical Watson-Crick pair with the incoming purine^{19,21,60,92}. This position is always uridine in adenine riboswitches and always cytosine in guanine/2'-deoxyguanosine riboswitches (Figure 1.6A). Similarly, the Watson-Crick face of nucleotide 51 is used to discriminate between nucleobases (adenine and guanine) and nucleosides⁹⁴. Residue 51 is always uridine in guanine and adenine riboswitches and cytosine in 2'-deoxyguanosine riboswitches (Figure 1.6B). This cytosine shifts toward nucleotide 74 such that it opens up the binding pocket to allow for the presence of the 2'-deoxyribose sugar moiety⁹⁴. Thus, small sequence changes to the aptamer can generate riboswitches that respond to chemically distinct effectors.

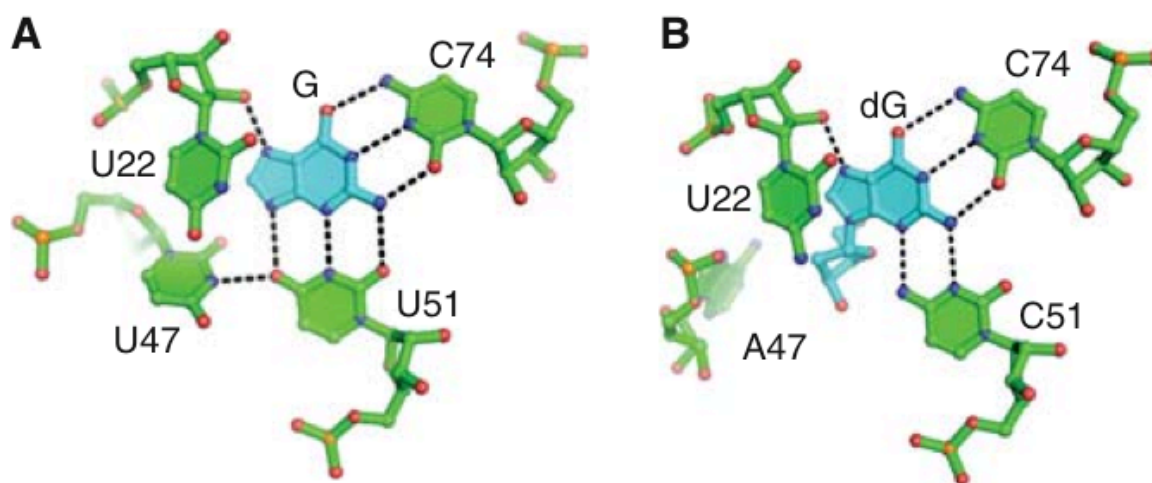


Figure 1.6 Examples of riboswitch specificity for purines

A) Recognition of guanine (cyan) by a pocket of conserved pyrimidine residues (green). The ligand interacts with the RNA primarily through hydrogen bonding interactions (black dashed lines) between pyrimidine 51 and 74. The binding pocket for adenine-binding riboswitches is nearly identical except that the ligand interacts with a uridine residue at position 74. (B) 2'-deoxyguanosine (cyan) is recognized through interactions with two cytosine residues (green). Note the shift of C51 relative to U51 to accommodate the 2'-deoxyribose moiety

A distinct challenge faced by RNA is recognition of negatively charged moieties. Instead of directly contacting negative charges presented by ligands, many aptamers recruit metal cations to help neutralize unfavorable electrostatic interactions between the ligand and RNA. For example, the phosphate groups of TPP and flavin mononucleotide (FMN) are coordinated by Mg^{2+} ions to mediate RNA contacts with the ligand^{59,63,93}. Similarly, as will be discussed in the following chapters, the carboxylic acid moiety of lysine interacts with the RNA via a potassium cation⁹⁵. The SAM-I and SAM-II riboswitches represent an exception in which the carboxylic acid moiety interacts directly with the Watson-Crick face of a guanine or adenine residue respectively^{69,75}.

Collectively these structures have highlighted what appears to be a general feature of the majority of riboswitch aptamer-ligand interactions: the ligand is highly

solvent inaccessible, buried within the core of the RNA (Figure 1.7). On average, 90% of the metabolite surface area is solvent inaccessible, which is greater than that of ligands bound to artificial aptamers (71%)⁹⁶. At first glance, this might be attributed to the need to discriminate between closely related metabolites in the cell¹¹¹. However, *in vitro* selected RNAs can easily achieve comparable specificities for their ligands without the need for extensive burial. The theophylline aptamer is capable of binding its cognate purine with an affinity that rivals the adenine riboswitch, capable of discriminating against caffeine by 10,000-fold⁹⁷. Like many riboswitches, the theophylline aptamer undergoes a significant ligand-dependent conformational change⁹⁸, indicating that folding and a high degree of ligand burial are not necessarily correlated phenomena.

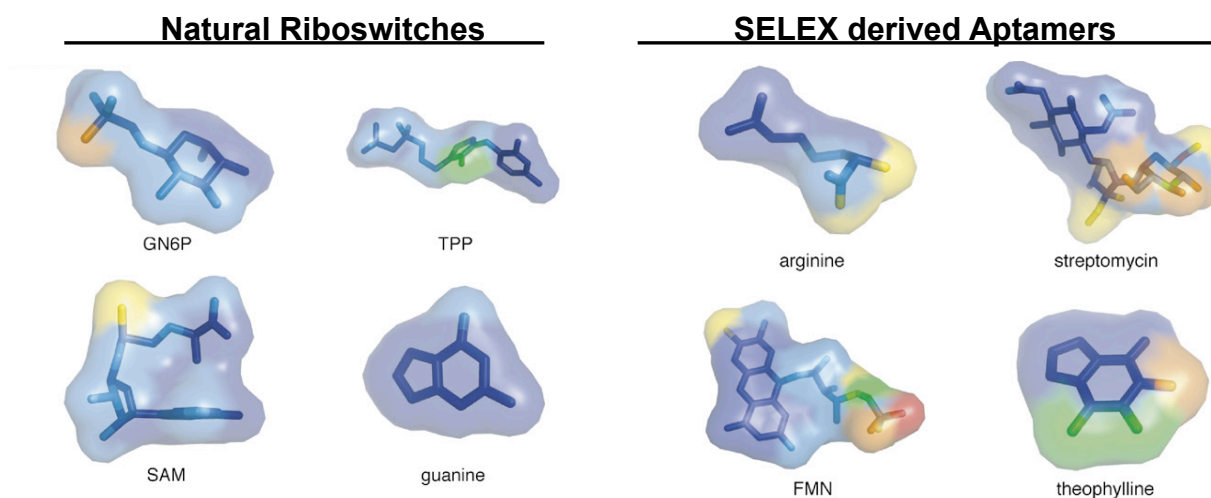


Figure 1.7 Comparison of solvent inaccessible surface area for ligands bound to naturally or artificially derived aptamers

The buried surface area of each ligand is indicated in blue, while solvent exposed portions are highlighted in yellow. Calculations were performed using a 1.4 Å radius probe as described in Edwards *et. al.* Curr. Opin. Struc. Biol. 2007. This figure is adapted from the reference.

1.5 Conformational changes associated with ligand binding

A more likely reason for extensive ligand burial is related to interdomain communication in the riboswitch. Solvent inaccessibility of the effector molecule implies that the aptamer must adopt a flexible “open” state that allows the ligand entry into the pocket, followed by a conformational change that encapsulates the ligand. This phenomenon, referred to as an induced-fit binding mechanism, is a common feature of RNA binding reactions, such as in RNP assembly⁹⁹. Conformational changes in the riboswitch aptamer domain are used to couple ligand binding to folding of downstream structural switch in the expression platform that in turn instructs the transcriptional or translational machinery.

The inherent flexibility of riboswitches in the absence of a ligand typically precludes crystallographic examination of these states. Instead, they are studied in solution using techniques such as chemical probing and NMR spectroscopy, which provide information on conformational dynamics at the global and local levels. NMR studies of both the *pbuE* and *xpt-pbuX* purine aptamers show that in the absence of ligand, nucleotides in the three-way junction are completely disordered even at ambient temperatures as indicated by the absence of peaks for corresponding nucleotides^{100,101}. This is consistent with the high degree of reactivity in these regions identified by in-line probing¹⁹ and SHAPE chemistry⁸⁰. Both NMR and chemical probing data indicate the J1/2 and J2/3 regions of the aptamer domain are disordered in the absence of the effector^{80,101,102}, suggesting that these elements form a flexible lid that closes around the docked ligand (Figure 1.5A). The loop-loop interaction on the other hand appears to be pre-established under physiological conditions^{80,83}.

Addition of the appropriate purine nucleobase induces structural organization of the three-way junction, as evidenced both by the appearance of new peaks in NMR spectra¹⁰³ and a 15°C increase in the T_m of these nucleotides by NMIA probing (Figure 1.5B)⁸⁰. Along with formation of intermolecular interactions between the ligand and RNA, these data reveal formation of additional intramolecular RNA interactions in the junction. Importantly, a subset of these ligand-dependent interactions involves formation of two base triples between J2/3 and the 3'-side of the P1 helix⁸⁰. Formation of these interactions was proposed to stabilize incorporation of the switching sequence into the receptor domain, serving as the basis for the communication of binding to the signal transduction domain⁶⁰. This was further supported by force spectroscopy of the pbuE aptamer, in which an additional ~2.2 pN of force was required to make the P1 helix bistable in the presence of adenine, providing direct evidence for adenine induced stabilization of this helix (Figure 1.5B)⁸⁷.

Although the purine aptamer appears to undergo localized conformational changes on binding ligand, others show more dramatic structural reorganization. For example, studies of the TPP riboswitch using site-specific labeling with the fluorescent reporter 2-aminopurine (2AP) revealed both the ligand binding pocket and a terminal loop-helix interaction fold with comparable rates upon the addition of ligand¹⁰⁴. However, a sevenfold disparity between the folding rate of the loop-helix interaction and the top of the P1 helix was measured, indicating that the free state of the TPP aptamer is largely disordered and ligand-induced organization of the RNA propagates from the periphery to the central junction. Similarly, fluorescence and NMR spectroscopic analyses of the preQ1 (a biosynthetic precursor of the queuosine nucleoside found in the anticodon of

many tRNAs) riboswitch revealed that the free state of this RNA folds into a hairpin structure with a highly disordered single stranded 3'-tail^{70,71}. Addition of preQ1 is required for association of the hairpin loop with the 3'-tail to complete the pseudoknot structure. The structural rearrangements associated with ligand binding therefore appear to be somewhat idiosyncratic, based largely on the topological arrangements of the RNA¹⁰⁵. It has been suggested that this feature of RNA folding is encoded by the nucleotides at key junction regions of structured RNAs¹⁰⁶.

1.6 Temporal considerations for transcriptional regulation

Many riboswitch aptamers have been reported to bind their effector ligands with equilibrium dissociation constants (K_D 's) in the nanomolar range. However, a number of studies have shown higher ligand concentrations are required to effectively regulate gene expression. The FMN, SAM-I, and lysine riboswitches for example have been reported to regulate require ligand concentrations ~100-1000-fold above the K_D of the aptamer to achieve half maximal transcription termination (T_{50})¹⁰⁷⁻¹⁰⁹. This is a hallmark of a kinetically controlled process in which the aptamer domain has insufficient time to equilibrate with the cellular environment before the expression platform commits the RNA to an alternative folding route that may be largely irreversible. Studies of a number of riboswitches have found that transcripts encompassing the entire riboswitch are deficient in effector binding, strongly arguing for irreversible folding^{27,83,110}.

The time sensitivity of riboswitch regulation can be explained in part by the slow kinetics of ligand binding, an inherent feature of induced fit processes. Measurement of the association kinetics of 2-aminopurine to the *xpt* and *pbuE* purine riboswitches

yielded rate constants on the order of $10^5 \text{ M}^{-1}\text{s}^{-1}$ for both aptamers^{111,112}. Similar values were obtained for the FMN and TPP riboswitches^{104,109}. These rates suggest that at 1 mM intracellular ligand concentration, at least 10 s would be required for effector to fully saturate the aptamer. If transcription occurred at a constant rate for this duration, a bacterial polymerase would transcribe nearly 500-1000 nucleotides, well beyond the boundaries of even the largest riboswitches, making regulation by the riboswitch impossible.

One means by which riboswitches deal with this time limitation is by manipulating the rate of transcription using programmed pause sites embedded within the expression platform. Two such sites were identified by synchronized transcription assays of the FMN aptamer that have pause lifetimes of 1 and 10 s respectively¹⁰⁹. Mutations that ablate pausing in these sites cause significant elevation of the T_{50} values in the context of the same aptamer, as expected when transcription proceeds through the riboswitch more rapidly. Uridine-rich tracts in the expression platform of the pbuE riboswitch have also been suggested to serve as transcriptional pause sites, though they remain to be experimentally validated¹¹². Although the general importance of pausing remains to be addressed for the majority of aptamers, this phenomenon may provide an important mechanism for tuning the response range of a riboswitch to ligand concentrations relevant for the cell.

Pausing may also provide time at important points during transcription to allow structural rearrangements that guide more efficient folding of the downstream sequence. For example, RNAs such as RNaseP, tmRNA, and SRP RNAs all form labile, nonnative structural intermediates at the point of transcriptional pauses¹¹³. These

intermediates sequester upstream portions of long-range helices in which the 5' and 3' sides are separated in primary sequence by more than 50 nucleotides. It has been proposed that these intermediates provide the RNA with a mechanism for preventing misfolding and enhancing the rate of the overall folding reaction *in vivo*. By analogy, the long-range P1 helix in riboswitch aptamers may be formed as part of a labile folding intermediate that is trapped by ligand binding. This would be consistent with the models offered by force extension studies discussed above (Figure 1.5B)⁸⁷.

1.7 Synthetic RNAs as a model for secondary structural switching

Riboswitches, as well as a number of other types of RNA regulatory elements in bacteria and eukarya, use mutually exclusive secondary structures to direct expression machinery. As discussed above it is important to consider the rate at which RNA transcribes and folds. It is also important to consider the relative thermodynamic stability of competing secondary structures. As opposed to the majority of studies that observe folding of a fully synthesized RNA, the directionality of transcription imprints order that biases the final conformation because upstream sequences can fold before transcription of downstream sequences. For riboswitches, this means that folding of the aptamer can occur in the absence of the expression platform, and the antiterminator can form before synthesis of the terminator stem-loop.

The effects of transcription order have been studied for many RNAs, but perhaps most clearly so using a simple bistable switch that mimics many of the features of naturally occurring riboswitches (Figure 1.8)¹¹⁴. This RNA was engineered to fold into mutually exclusive branched or rodlike structures with free energy structures that were predicted to be equivalent. Transcription of the sequence in a forward direction (A→D,

Figure 1.7A) yields exclusively the branched structure as determined by native gel electrophoresis, whereas transcription of exact same sequence in the opposite direction (D→A) leads to the alternative rodlike structure with ~90% frequency. In contrast, heat renaturation of the RNA leads to an equal distribution of the two states, which do not significantly interchange at room temperature. In other words, RNA can readily adopt non-equilibrium structures during cotranscriptional folding because of high activation energies associated with their interconversion¹¹⁵. This result emphasizes that traditional methods for studying RNA folding might fail to capture some of the most crucial features of cotranscriptional folding that are inherently tied to the mechanism of riboswitch regulation.

The analogy between the engineered and natural riboswitches can be extended by considering self-induced folding transitions occurring during folding of the reverse sequence (Figure 1.8B). In the absence of a competitor oligonucleotide, the reverse switch sequence destabilizes the P_b stem in favor of the P_c stem; lower temperatures bias the structure further toward P_c because of slower transcription rates. This refolding process can be disrupted by introduction of a competitor oligonucleotide (Figure 1.8B, black line) that pairs with the P_c stem early in transcription. The oligonucleotide efficiently redirects folding into the branched structure, in the same fashion as ligand binding prevents self-induced transitions during riboswitch folding that would lead to aberrant antitermination (Figure 1.8B). Elements of the unbound aptamer can be thought of as behaving like the antisense oligonucleotide by disrupting terminator stem formation while simultaneously nucleating the antiterminator stem (Figure 1.8C). This comparison warrants further investigation of the simple secondary structures in both

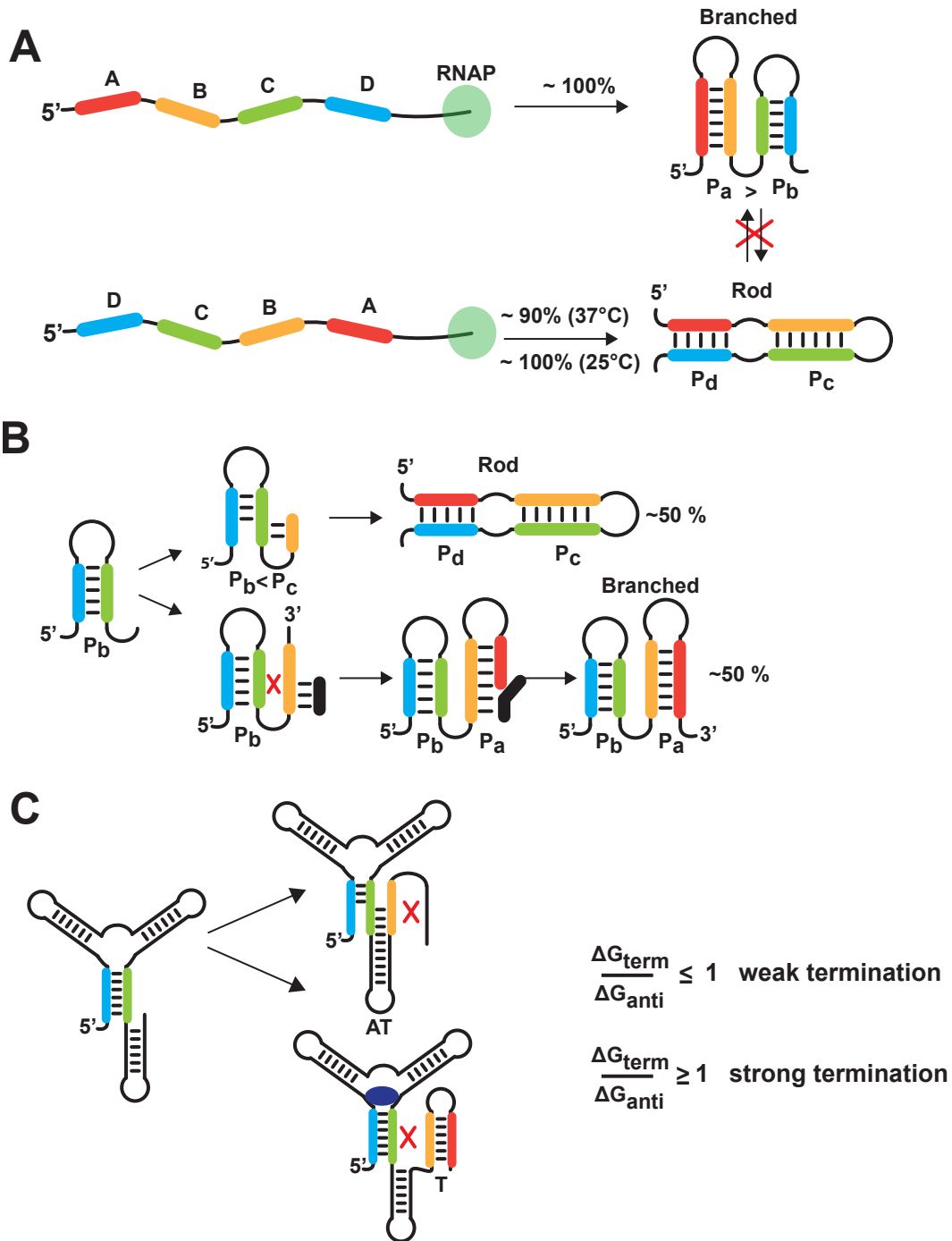


Figure 1.8 Models for cotranscriptional folding of synthetic and natural RNA switches.

(A) A synthetically designed RNA sequence transcribed in forward and reverse directions illustrating the influence of transcription order on the folding outcome. The free energy of each sequential helix dictates the outcome of folding. (B) A model of the reverse synthetic sequence reveals that structural transitions can be redirected by addition of a competitor oligonucleotide (black) to the transcription reaction. (C) Regulatory efficiency of purine riboswitches correlates with the relative free energy of the terminator (T) and anti terminator (AT) stems, pointing to similarities between riboswitches and their synthetic counterparts.

domains of the RNA that must interchange during transcription to elicit regulatory control.

Another key conclusion gained from studies of the engineered switch is that the relative stability of helices ($P_a > P_b > P_c > P_d$) is sufficient to encode a precise folding pathway¹¹⁴. Similar models have been devised to explain the variable regulatory activities of purine riboswitches within the *B. subtilis* genome¹¹⁶. In cases where the relative stability of the antiterminator hairpin (ΔG_{anti}) is greater than that of the terminator hairpin (ΔG_{term}) (Figure 1.8C), reporter gene assays demonstrate that there is a significant amount of transcription read through *in vivo*, even at saturating ligand concentrations in the growth media. In contrast, increasing the stability of the terminator hairpin is correlated with the enhanced ability of the riboswitch to promote transcription termination under high ligand concentrations¹¹⁶. Similar factors may account for the large range of behaviors observed for the 11 SAM-I riboswitches in the *B. subtilis* genome¹⁰⁸. The variability of expression platforms in nature suggests that the secondary structural elements have been tuned to provide an optimal folding pathway based on the particular genetic context of the riboswitch. Better understanding of this process should therefore lend functional insight into this variation, and enable reverse engineering of expression platforms to construct novel biosensors with predictable regulatory behaviors.

1.8- Regulation of lysine metabolism by riboswitches

My research has focused in particular on one of the earliest identified classes of riboswitch, the lysine riboswitch. This RNA is found primarily in the firmicutes and

fusobacteria, and represents one of the most abundant riboswitch classes thus far discovered¹¹⁷(Figure 1.1). Lysine riboswitches have been found upstream of a variety of different genes that are involved in the biosynthesis, utilization and transport of this essential amino acid^{23,118}. Most are however associated with lysine biosynthesis, which typically occurs via the diaminopimelate (DAP) pathway in bacteria and some species of plant (Figure 1.8). This pathway utilizes aspartate to generate a number of compounds in addition to lysine that are essential to normal bacterial growth and sporulation. For example, DAP is an important component involved in the cell wall biosynthesis, and dihydropicolinate makes up a large fraction of the dry cell weight of sporulating bacteria such as *Bacillus anthracis*¹¹⁹. The phosphorylation of aspartate (aspartate 4-phosphate) is also a crucial step in production of other essential amino acids such as serine, threonine and methionine (Figure 1.9)

The myriad roles of the diaminopimelate pathway in microbial physiology has attracted interest in targeting this pathway with novel antimicrobials, particularly as homologous pathways are absent in mammals¹²⁰⁻¹²². Indeed many compounds have already been identified that inhibit key enzymes such as the dihydropicolinate synthase in *Mycobacterium tuberculosis*¹²² and DAP analogs that inhibit the growth of various other bacterial strains¹²⁰. For historical reasons the antimicrobial lysine analog S-2-aminoethylcysteine (AEC Figure 1.10 box) is of particular relevance to my work on the lysine riboswitch. Efforts to elucidate the mechanisms of AEC toxicity in *E. coli* and *B. subtilis* uncovered mutations that clustered in the leader sequence of the *lysC* operon that encodes the subunits of the aspartokinase responsible for aspartate phosphorylation^{123,124}. Expression of this operon is strongly repressed by excess lysine,

and mutations to the leader sequence of the transcript were found to release this regulatory repression, leading to constitutive lysine production¹²³. Comparison of the leader sequences from these two organisms suggested a common structural element, however the mechanism of its function remained unclear until more genomic sequences became available.

Comparative analyses of the *lysC* leader sequences from numerous bacterial genomes have now revealed that the leader sequence folds into a structure that consists of five paired regions (labeled P1-P5, Figure 1.10A) whose specific length restrictions indicate their involvement maintenance of a complicated structure^{22,23}. The presence of a long-range kissing-loop was inferred from covariation between the L2 and L3 elements within the multiple sequence alignments²², and the recurrent kink-turn⁵⁰ and loopE⁵¹ motifs can be readily identified within long P2 helix. At the center of these helices is the five-way junction element that is highly conserved across all phylogeny. Interestingly this structural model implies that a relatively complex architecture is required to confer specificity for a simple amino acid. In-line probing studies indeed confirmed that the RNA achieves specificity for L-lysine over a number of closely related compounds, including various intermediates of the DAP biosynthesis pathway, and the stereoisomer D-lysine²². On the other hand, when challenged with compounds that present bulky additions to certain positions in the alkyl chain or at the side chain terminus of the amino acid the RNA was found to bind with relatively high affinity⁸⁹ (Figure 1.10B), raising interesting questions about the nature of the ligand binding site. These studies also identified additional antimicrobial agents whose toxicity, like AEC, could be traced in part to the regulatory activity of the lysine riboswitch, suggesting the

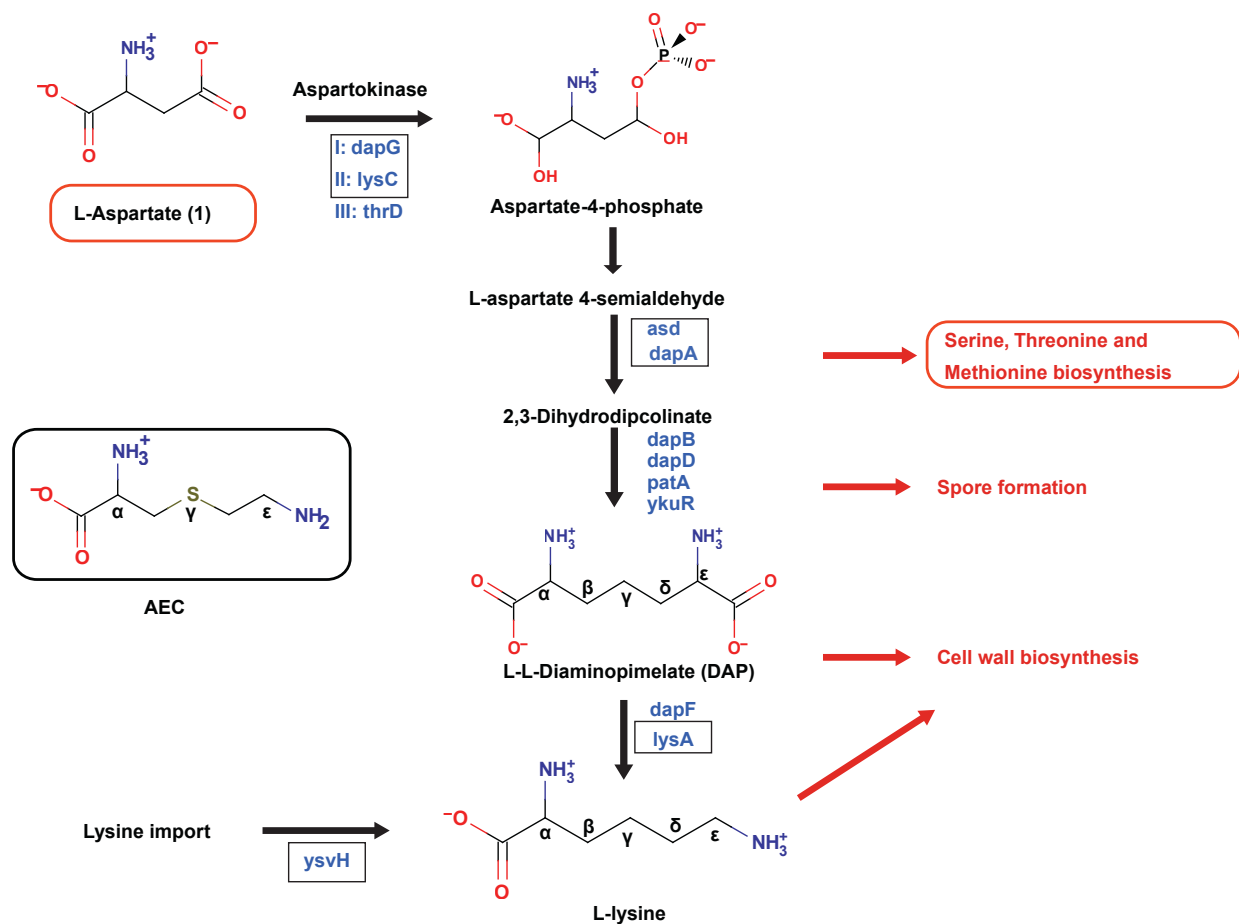


Figure 1.9 Diaminopimelate pathway for lysine biosynthesis in bacteria.

Aspartate is converted through a series of enzymatic steps into L-lysine by the genes indicated in blue. Genes frequently associated with lysine riboswitches are indicated by the boxes. The red arrows indicate that in addition to protein synthesis, lysine and many key intermediates from this pathway are crucial for cell wall biosynthesis, sporulation, and the synthesis of other essential amino acids. Equivalent carbon positions of AEC, DAP and lysine are labeled for clarity.

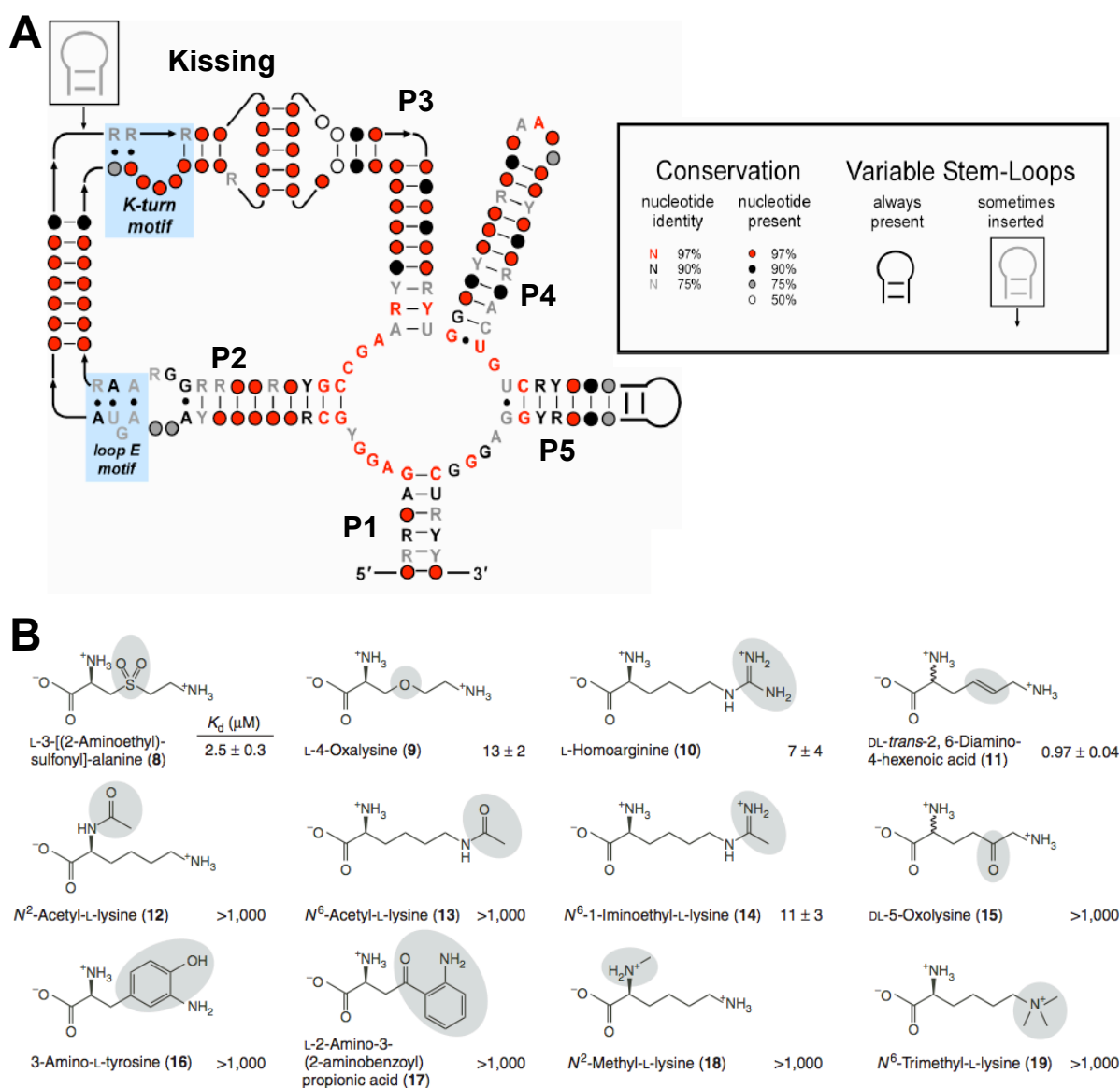


Figure 1.10 Consensus structure and specificity profile of the lysine riboswitch

A) Secondary structural map of the lysine riboswitch implicates the 5-way junction in ligand binding. Nucleotide conservation patterns are defined as indicated below. Figure taken from Barrick *et al.* 2007 Genome Biol. B) Binding studies of lysine analogs demonstrate that the alkyl chain is tolerant to substitutions. Some of these compounds such as the L-3-[(2-Aminoethyl)-sulfonyl]-alanine and L-4-Oxalysine were also demonstrated to have antimicrobial activity in liquid culture. Figure taken from Blount *et al.* 2007.

potential of this element as an antimicrobial target⁸⁹.

Interestingly, early studies of the lysine riboswitch not only suggested the antimicrobial potential of this RNA, but also implied that the riboswitch is under kinetic control; in other words the ligand concentrations required to elicit a regulatory response during transcription are in vast excess of the observed $K_D^{22,107}$. Such observations suggest that the kinetics of folding and ligand binding play an important role in determining the outcome of the regulatory response. These studies were however poorly controlled and failed to properly account for some of the important variables that influence the apparent binding affinity ($K_{D,app}$) such as the presence of specific cation species and temperature making it difficult to differentiate between kinetic control and condition dependence of the $K_{D,app}$.

To address aspects of both ligand specificity and transcriptional regulation by the lysine riboswitch I employed a combination of biophysical studies of the aptamer domain and functional studies of transcriptional regulation by the fully intact riboswitch to develop a more accurate picture of ligand binding and regulation. The following chapters detail the structural models of the lysine riboswitch aptamer domain obtained by X-ray crystallography and complementary solution based techniques such as chemical probing and small angle X-ray scattering techniques that were utilized to address the structural rearrangements that accompany ligand binding. These studies collectively aimed to provide mechanistic insight into the features are presumably important in directing folding of the downstream secondary structural switch.

Biochemical experiments were also conducted to examine particular aspects of molecular recognition by the aptamer domain using a variety of lysine analogs that

present alterations to key contacts at the RNA-ligand interface in order to determine their influence on the thermodynamics and kinetics of ligand binding. This data was used to derive expected regulatory responses for the various ligands based on a simple kinetic model of regulation and compared to empirical observations of regulation obtained using a minimally reconstituted *in vitro* transcription assay to determine the accuracy of the kinetic model and gain further insight into the regulatory landscape of the lysine riboswitch.

Lastly, I will describe experiments that were undertaken to define the mode by which the aptamer domain and expression platforms are able to communicate the status of ligand occupancy between one another to ensure a high fidelity regulatory response. This was accomplished by a variety of approaches that included a systematic mutational analysis to the P1 helix of the RNA as this was hypothesized to be the key element of allosteric communication. Indeed, it was demonstrated that stability of the P1 helix is the primary determinant in the folding pathway chosen by the expression platform, and that optimal stability of P1 is essential to ligand binding to folding of the secondary structural switch. This insight aided in the design and optimization of chimeric riboswitches that take advantage of the apparent modularity of both the aptamer domain and expression platform. This exciting approach was found to provide a robust and reproducible means by which RNA aptamers can be incorporated into the context of a regulatory device.

Chapter 2: Structural basis for recognition of lysine by mRNA.

As described in the previous chapter, comparative genomics analyses identified a conserved structural element upstream of a variety of genes involved in lysine biosynthesis and transport^{22,23}. This element is composed of a set of five helical elements that center on a highly conserved five-way junction, in which a large number of residues are invariant across disparate bacterial phyla. As a first step toward understanding how this RNA achieves lysine specificity I sought to develop atomic resolution models of the riboswitch aptamer bound to lysine. The following chapter details this work along with the complementary solution based techniques that were employed to study the conformational changes associated with ligand binding and regulation.

2.1- Crystallization and structural determination of the *asd* aptamer from *Thermatoga maritima*

In order to understand the basis for lysine recognition and AEC resistance, I sought to crystallize the RNA bound to its cognate ligand. To achieve this goal I designed a library of constructs for aptamers from the *T. maritima*, *B. subtilis*, *Haemophilus influenzae* and *Shigella flexneri* genomes. These libraries were constructed by making systematic changes to the length of the P1 and P5 helices (Figure 2.1). This strategy requires variations to be engineered in non-conserved helical structures and has proven highly robust for rapidly obtaining engineered RNA constructs that are suitable for high resolution structural determination¹²⁵. Individual library members were subjected to a sparse matrix screening using hanging drop vapor diffusion. Screens were performed using commercially available kits including the

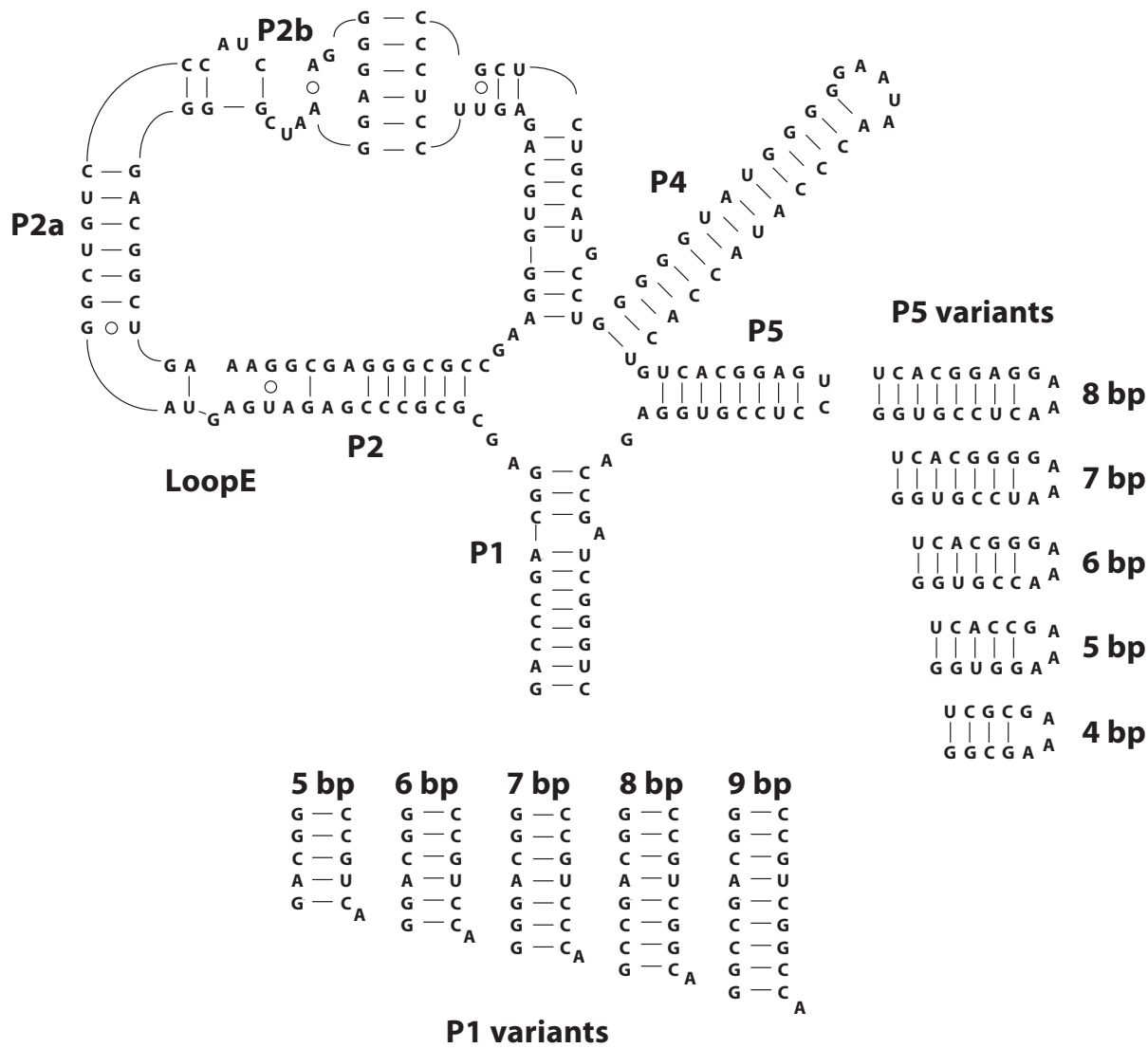


Figure 2.1 Crystal construct design strategy for the *T. maritima asd* aptamer

Constructs containing P1 stems ranging from 5-9 bp were combined with different variants of P5 truncations and screened individually against sparse matrix conditions. The organismal name, P1 length and P5 length provide a simple nomenclature for identification. For example the *T. maritima* aptamer with a P1 helix of 6 and a P5 helix of 4 is designated as the *Tma64* variant. The same design principles were also applied to the other phylogenetic variants

Natrix, Nucleic Acids mini screen, and Crystal Screens 1 and 2 (Hampton research) as these provide a broad survey of different buffer, pH, ionic strength, precipitant and analyte conditions that promote macromolecular crystal formation.

Sparse matrix screening revealed a set of conditions (2 M $(\text{NH}_4)_2\text{SO}_4$ or Li_2SO_4 with or without 5 mM MgCl_2) that supported crystallization of variants from many of the phylogenetic scaffolds (data not shown). These conditions allowed the crystals to be flash frozen without additional screening for cryoprotectant, and screened at the rotating copper anode X-ray source at CU Boulder. Initial screening and data collection efforts identified a variant of the *T. maritima* aptamer (*Tma64*, Figure 2.2A) that diffracted 2.8 Å resolution in the bound conformation and 2.95 Å in the unliganded state to yield the initial native data sets. To obtain phase information for these data sets, we attempted to co-crystallize the RNA in the presence of iridium hexamine, a strategy that proved successful in yielding a heavy atom derivative with sufficiently high quality (2.8 Å resolution) for phase determination using single-wavelength anomalous diffraction (SAD) experiments (Figure 2.1 C). The experimentally derived phases from the SAD data produced an electron density map of sufficient quality to enable iterative cycles of modeling and refinement using the Coot¹²⁶ and Phenix¹²⁶ software systems. The model was refined to a final $R_{\text{work}}/R_{\text{free}}$ of 18.2/20.9 (Table 2.1) suggesting that the model is a highly accurate representation of the RNA.

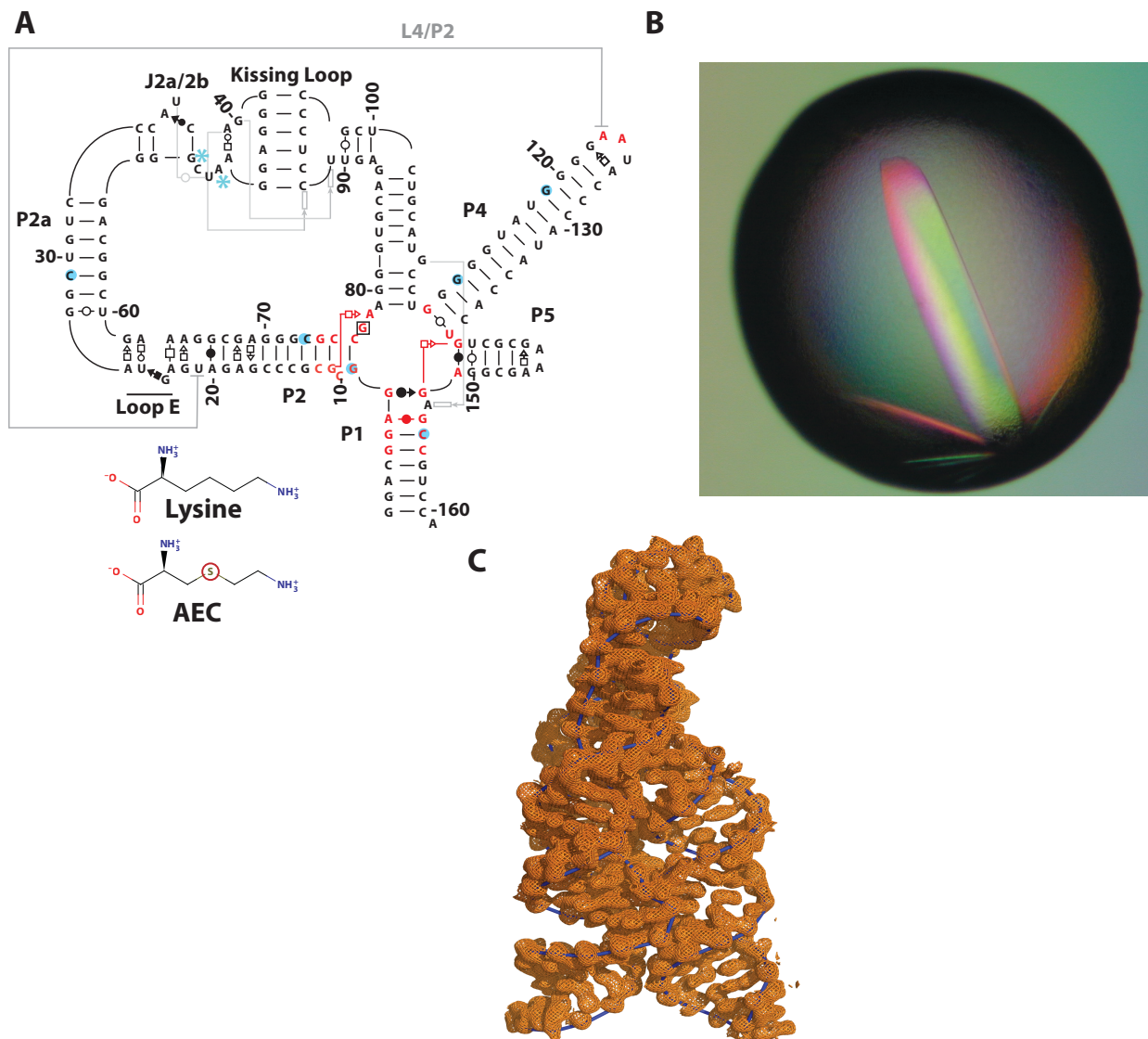


Figure 2.2 Structural determination of the lysine riboswitch.

A) The secondary structure of the lysine riboswitch aptamer construct derived from 5' leader sequence of the aspartate semialdehyde (*asd*) gene in *T. maritima*. Base-pairing interactions are shown using the nomenclature of Leontis and Westhof. Important secondary and tertiary structural motifs are labeled accordingly. Bold nucleotides shown in red are $\geq 90\%$ conserved across all phylogenetic variants, while those in black indicate positions whose presence is invariant, but have lower sequence conservation. Positions where mutations have been identified to confer resistance to AEC are circled in blue. Asterisks indicate mutations to the canonical kink-turn motif^{44,45} that is typical of most lysine aptamers but is replaced by the J2a/2b motif in this RNA. The molecular structures of lysine and AEC are shown for reference. B) Representative crystal of the lysine riboswitch used in structure determination. C) Global perspective of the initial electron density map of the lysine bound structure demonstrates a clear helical pattern that allowed for reliable modeling of the pertinent structure.

Table 2.1: Data Collection, phasing and refinement statistics

	Lysine-RNA complex Ir-hexammine derivative	Free RNA
Data collection		
Space group	P3 ₂	P3 ₂
Cell dimensions		
a, b, c (Å)	119.82, 119.82, 58.74	120.19, 120.19, 58.25
α , β , γ (°)	90, 90, 120	90, 90, 120
Wavelength (Å)	1.1050	1.5418
Resolution (Å)	40.0–2.8 (2.91 - 2.8)*	19.70–2.95 (3.06-2.95)
R_{sym} or R_{merge} (%)	8.4 (34.8)	9.0 (35.5)
$I/\sigma I$	17.9 (4.4)	10.2 (3.4)
Completeness (%)	99.5 (96.6)	99.5 (100)
Redundancy	5.2 (3.7)	3.62 (3.63)
Refinement		
Resolution (Å)	32.6–2.8 (2.87-2.8)	17.11-2.95 (3.02-2.95)
No. reflections	23986 (97.7%)	19610 (99.4%)
$R_{\text{work}}/R_{\text{free}}$	18.20/20.86	18.61/22.04
No. atoms	3631	3547
RNA	3491	3491
Ligand	10	N/A
Water	55	56
Atomic Displacement	54.61	54.92
Parameters (B-factors), all		
RNA	47.22	55.12
Ligand/ion	35.76	N/A
Water	39.14	42.87
r.m.s. deviations		
Bond lengths (Å)	0.005	0.004
Bond angles (°)	1.251	1.239
Maximum likelihood coordinate error (Å)	0.32	0.43

Data was collected from a single crystal.

*Highest resolution shell is shown in parenthesis.

2.2- Global Organization of the lysine riboswitch

The tertiary architecture of the lysine riboswitch is (Figure 2.3) is dominated by three coaxial stacks formed between the P1/P2 (dark blue), kissing-loop/P3 (dark blue/light blue) and P4/P5 (green) helices. These coaxial stacks are arranged in a parallel topology, similar to the global organization of other functional RNAs and riboswitch aptamers. This conformation is stabilized via key tertiary interactions mediated by the terminal loops of the P2, P3 and P4 helices. A kissing loop interaction is observed between the terminal loops of P2 and P3 (L2 and L3 respectively, Figure 2.3C and Figure 2.4A) that was identified as important for the ability of the *B. subtilis* lysC riboswitch to efficiently terminate transcription¹⁰⁷. Unlike other structurally characterized kissing loop interactions such as those that promote genome dimerization in the human immunodeficiency virus¹²⁷, there is a stacking interaction between G40 and U91, which are oriented perpendicular to the P2b/3 helical axis (Figure 2.4A). These two bases form hydrogen-bonding interactions with the major groove of the central four base pairs of the P2b/3 helix. This additional dinucleotide “staple” may constitute an adaptation for function at elevated physiological temperatures as similar observations were made in the selection of thermophilic ribozymes, where mutations that add new tertiary interactions or further stabilize existing ones are responsible for adaptation to elevated temperatures¹²⁸.

The ability of the two terminal loops to interact is achieved by a ~120° bend at J2a/2b using a novel internal loop motif (Figure 2.1 and 2.3C). In the majority of the phylogenetic variants of the lysine riboswitch this turn is achieved by the structurally similar canonical kink-turn motif⁵⁰. Thus, while the majority of the aptamer domain is

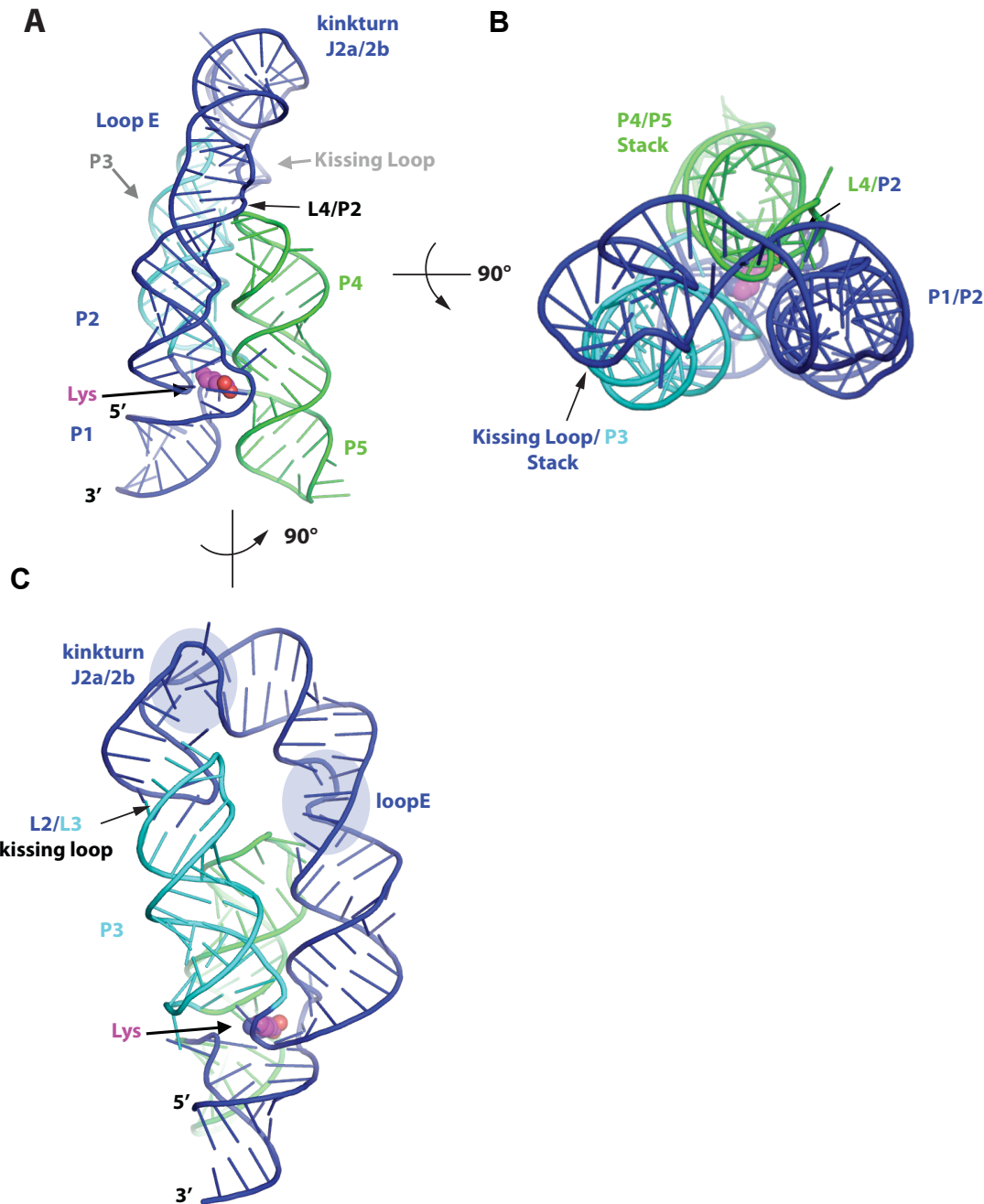


Figure 2.3 Global view of lysine riboswitch tertiary structure.

A) Helices are colored to highlight the three coaxial stacks that characterize the global architecture of the RNA. Lysine is represented as van der Waals spheres, buried in the center of the 5 way junction. B) A 90° rotation about the x-axis of perspective shown in A) reveals that the global structure consists of a three helix bundle, in which the P4/P5 coaxial stack is packed tightly against the P1/P2 and L2/P3. This structure is stabilized by the long-range L4/P2 and L2/L3 kissing loop interactions. C) 90° rotation about the z-axis of perspective in A) allows visualization of the structural motifs that enable folding of this aptamer. The loop E motif extends and introduces a ~ 50° bend in the P2 helix and the kink-turn like motif bends the helix another 130° to form the kissing loop interaction.

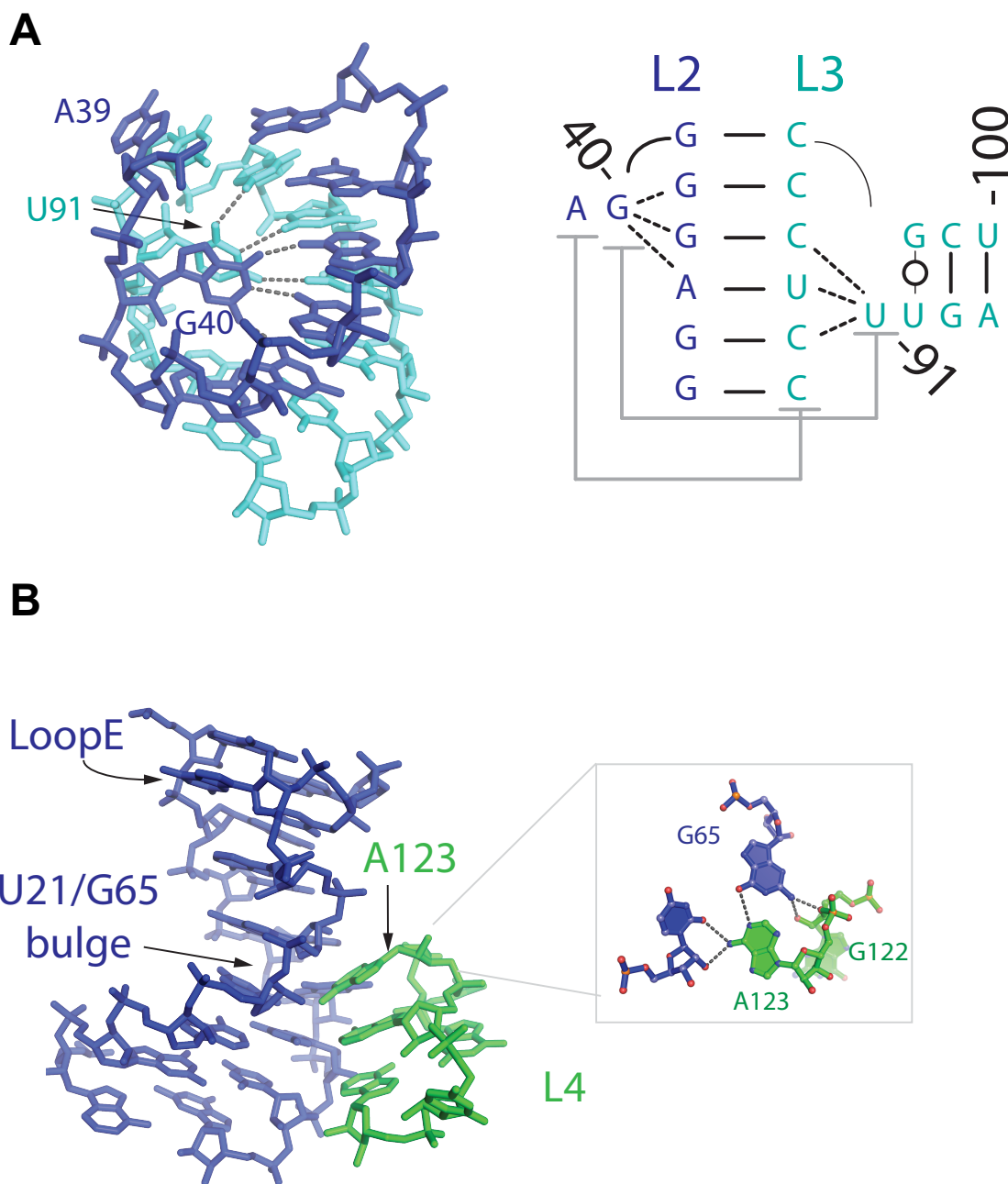


Figure 2.4 Structural motifs that stabilize global architecture of the lysine riboswitch

A) The unusual kissing loop complex of the lysine riboswitch is hyperstabilized by a pair of nucleotides (U91 and G40) that form H-bonding interactions with the major groove face of the long-range helix. A39 Stacks on top of the terminal G-C base pair of the kissing loop to provide additional stability. B) The terminal loop of P4 (L4) docks forms an unusual A-minor like interaction at a symmetrical bulge in P2. The primary contacts are shown to the right, in which the universally conserved A123 nucleotide forms interactions with the sugar face of U21 and the Watson Crick face of G65.

highly conserved, some elements of the peripheral region of the lysine riboswitch have evolved unique solutions to the stabilization of a common global architecture, reflecting the modular nature of RNA structure in general¹²⁹.

The most highly conserved element involved in stabilizing the three-helix bundle is an interaction between the terminal pentaloop of P4 and an internal loop motif that is adjacent the consensus sarcin/ricin or loopE motif^{51,130} between P2 and P2a (Figure 2.4B). The pentaloop of P4 forms a structure similar to a standard GNRA tetraloop motif by flipping out a uridine residue (U125)¹³⁰. Rather than docking with another helix using the sugar edge of the three stacked adenosine residues, as commonly observed for most tetraloop-mediated interactions⁵⁴, the adenine bases interact with the minor groove of P2 using their Watson-Crick faces. Unusually, A123 forms the central base of a U21-A123-G65 base triple that anchors the interaction.

2.3- Lysine recognition

The ligand-binding pocket is contained within the core of the five-way junction motif, sitting between the P1 helix and J2/3 and is flanked by the first base pairs of the P2 and P4 helices (Figure 2.4). The carboxylate group of lysine forms a set of hydrogen bonds with the N2 amino groups of the G111•U137 wobble pair, the G9-C76 Watson-Crick pair and the 2'-hydroxyl group of G8. An independently published structure of this RNA also identified the role of a potassium ion (K^+) in neutralizing electrostatic repulsion between the negatively charged carboxylate and the lone pair of electrons on the N3 of the G9 nucleotide⁹⁵. While our structures were solved in the absence of K^+ suggest that this ion is not absolutely essential for binding, biochemical data demonstrated that K^+

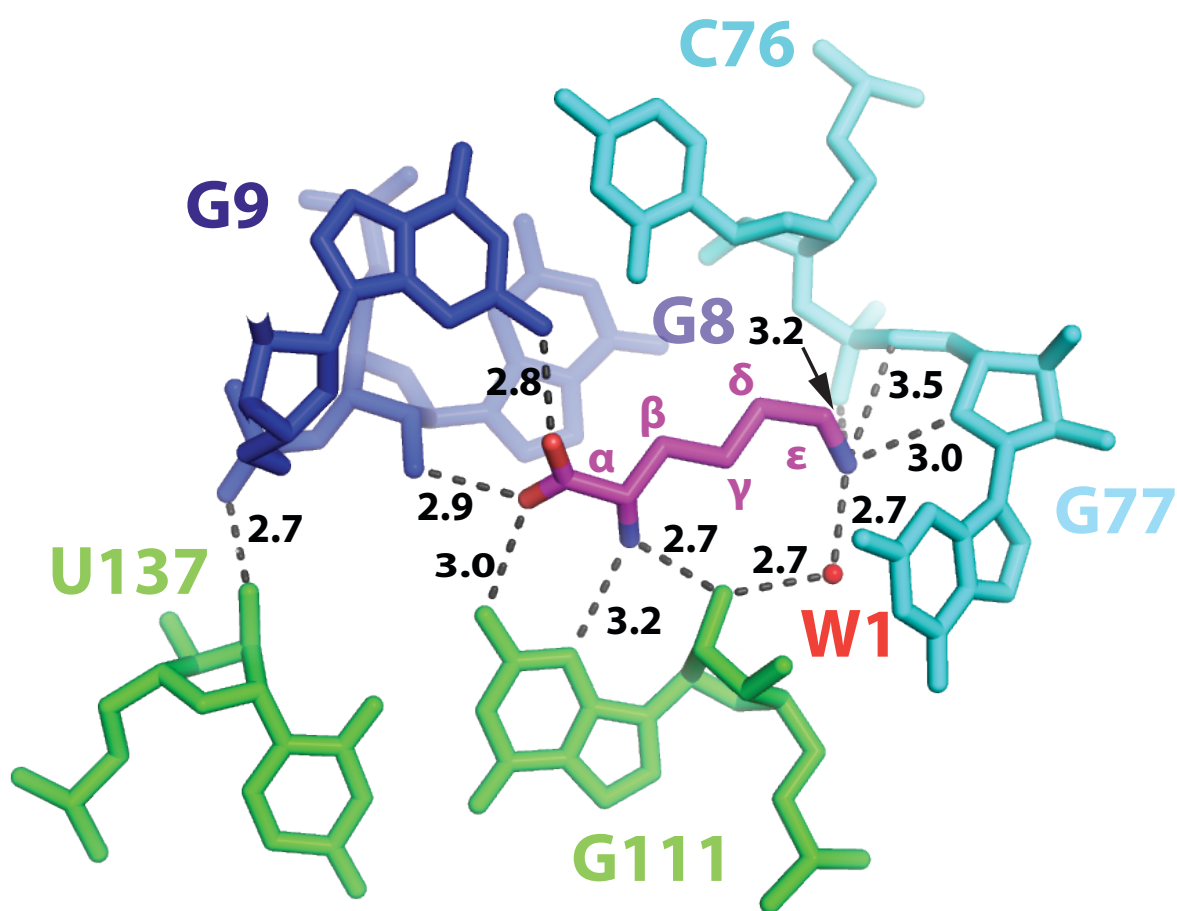


Figure 2.5 Details of lysine recognition.

The distances of the bonds between lysine and RNA are given in Å. The α through ϵ carbons are labeled accordingly. The γ -position which is tolerant to a variety of substitutions (see text) is indicated by the oval (magenta).

enhances the binding affinity of the RNA for lysine by 10-50 fold⁹⁵. Further contacts to the lysine main chain involve hydrogen bonds between the N3 and O2' atoms of G111 to the α -amino group.

The ϵ -amino group that defines lysine from other amino acids is recognized by a combination of electrostatic and hydrogen bonding interactions within a pocket that places it close to the non-bridging phosphate oxygen of G77 along with the O4 oxygen atom of the ribose sugar. The relatively small size of the ϵ -amino pocket near G77 precludes efficient recognition of arginine and *N*⁶-trimethyl-L-lysine due to steric hinderance²². This is also likely the basis for discrimination between the related metabolites lysine and diaminopimelate, which contains an additional carboxylate at the δ position (Figure 1.8 and Figure 2.5). Interestingly we observed an ordered solvent molecule within H-bond distance from the ϵ -amine that appears to make an extended network of H-bond interactions with the 2'-OH groups of G111 and G152 (not shown). G152 is implicated in the formation of the mutually exclusive antiterminator structure, implying that this solvent mediated contact may contribute directly to regulatory control. The importance of solvent mediated contacts is supported by independent crystal structures that identified this water mediated interaction⁹⁵.

Discrimination between lysine and other closely related compounds is further achieved through indirect recognition of the methylene linker of the side chain. The lysine side chain is bound in an extended conformation that allows it to span the two sites of interaction of the polar atoms, consistent with the ability of a lysine analog that contains a trans-double bond between the γ - and δ -carbons to productively bind this

riboswitch⁸⁹(Figure 1.9B). As a result, compounds containing shorter or longer side chains (L-ornithine and L- α -homolysine, respectively) are not efficiently bound because their side chain is of the incorrect length to allow the proper contacts between all of the polar atoms of lysine and the RNA. The hydrophobic methylene groups are primarily contacted through stacking interactions with G77, A78, and the G8•G152 pair.

Despite being a critical component of proper recognition the methylene groups of lysine are not tightly packed against the RNA. The loose packing around lysine explains the ability of antimicrobial lysine analogs containing modifications at the γ -position (Figure 1.9 and Figure 2.5), such as L-3-(2-aminoethyl)-sulfonyl-alanine, L-4-oxalysine, and AEC, to bind reasonably well to the riboswitch^{22,89}. This feature of lysine recognition is similar to the TPP riboswitch in that the central thiazole ring of TPP is poorly recognized by the RNA⁵⁹. In both cases moieties recognized through indirect readout-which are generally the hydrophobic groups-are modified to yield riboswitch-binding antimicrobial agents⁴⁶.

2.4- Ligand-dependent conformational changes in the RNA

Lysine is completely buried within the five-way junction (100% solvent inaccessible), indicating that there is some form of folding event concurrent with binding. To test the hypothesis that lysine-dependent conformational changes occur within the riboswitch, I crystallized the RNA in the absence of lysine and determined its structure. Surprisingly, the RNA crystallized under the same conditions and in the same space group. The resulting structure is nearly identical to the complexed form (Figure 2.6), with only minor differences between the two structures particularly along the 5'-side of the

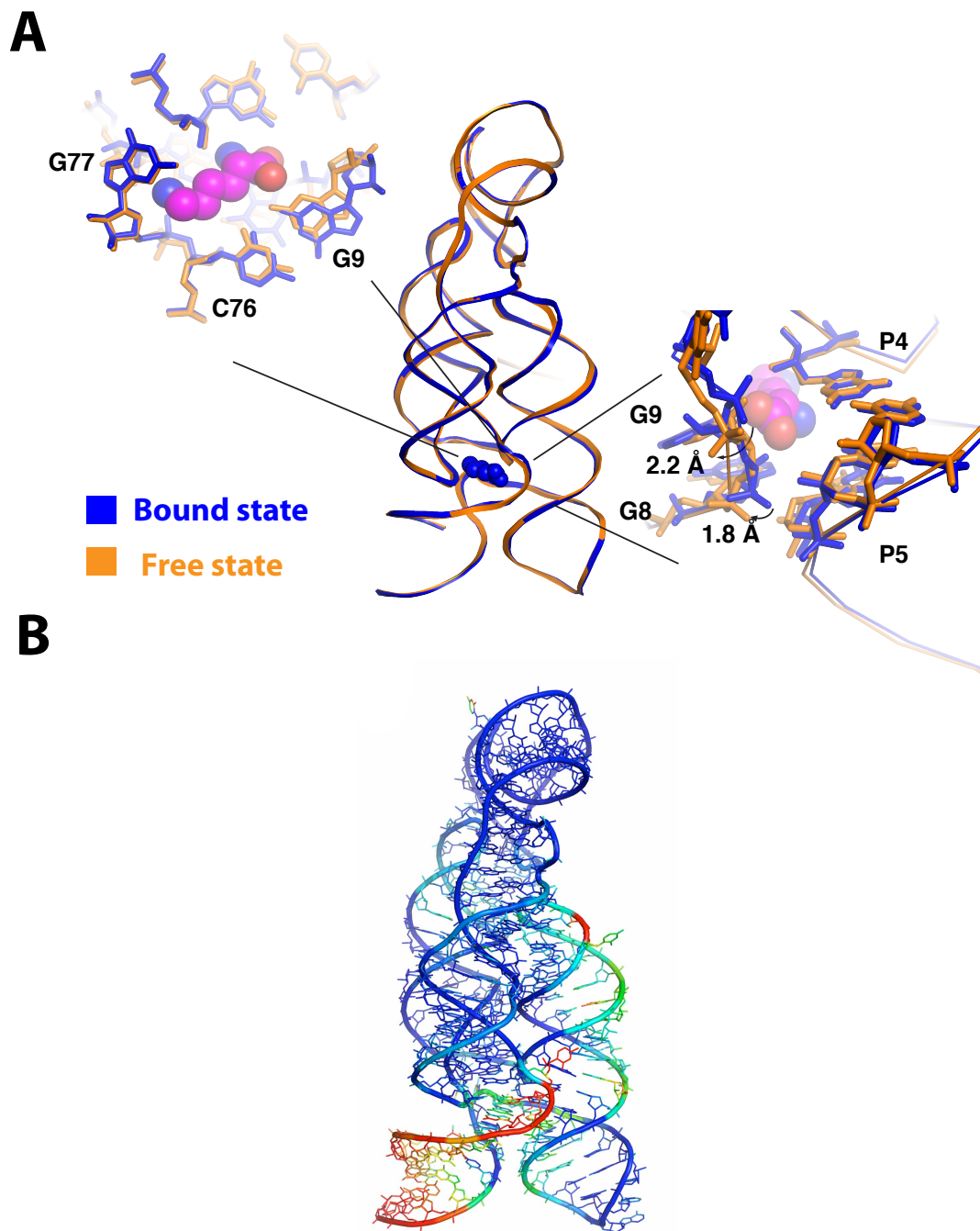


Figure 2.6 Comparison of bound and free state structures observed by X-ray crystallography.

A) Global superposition of the free (orange) and bound (blue) aptamer using the maximum likelihood theseus alignment program. The two structures superposition with a maximum likelihood r.m.s.d. of 0.08 Å (classical pairwise r.m.s.d is 0.70 Å). Close ups of the binding pocket show that there are small shifts in the binding pocket, particularly at position G9 which makes contacts with the carboxyl terminus of lysine. B) Map of the estimated variance between the atomic coordinates of the two structures; blue represents low variance (< 1 Å) and red denotes high variance (> 1 Å).

P1 helix (Figure 2.6B). Close examination of the binding pocket shows that positioning of some of the nucleotides is perturbed by ~ 2 Å, but the overall pattern of base interactions remains the same. This finding suggests that the global architecture can be formed in the absence of ligand. How then does the ligand gain access to the binding site? From the crystal structure it is clear that a solvent accessible channel exists near the carboxyl terminus of the binding site formed between the 5' side of the P1 helix particularly nucleotides G8 and G9) and the P4/P5 coaxial stack (Figure 2.5A, right). If the structure is largely preformed, then this channel appears to be the likely point of entry for the ligand⁹⁵. Consistent with this hypothesis the G8 and G9 nucleotides appear to swing out slightly which may act as a mechanism for gating accessibility to the binding site.

On the other hand, similarities between the bound and free conformations could potentially be an artifact of the crystal lattice whereby crystallization induces a conformation that does not exist in solution. To further explore the potential similarity between the bound and unbound lysine riboswitch in solution, I probed backbone flexibility using SHAPE chemistry⁸¹ to monitor local changes in the RNA and small angle X-ray scattering (SAXS)¹³¹ to monitor global conformational changes. To ensure that the observed behavior is a general feature of lysine riboswitch aptamers, I probed both the *T. maritima asd* and *B. subtilis lysC* RNAs at varying magnesium concentrations in the absence and presence of 670 μ M lysine (Figure 2.7 and Figure 2.8). In particular, I wished to probe these RNAs under near physiological salt concentrations (100 mM NaCl, 0.5–2 mM MgCl₂), as opposed to the high salt conditions used for crystallization (2 M Li₂SO₄, 5 mM MgCl₂, 60 mM iridium hexamine). These experiments reveal two

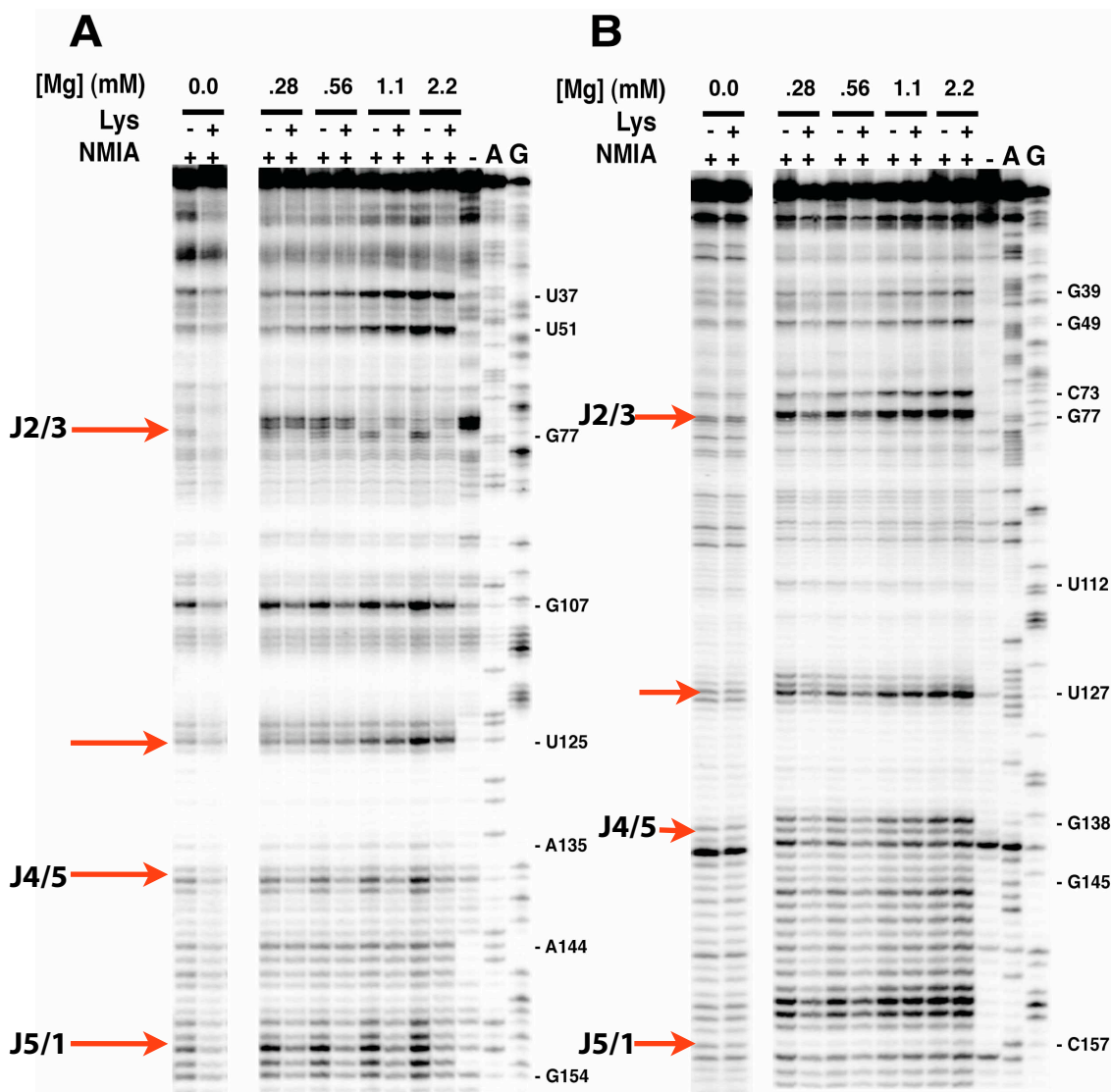


Figure 2.7 Conformational changes associated with Mg^{2+} and lysine binding
SHAPE probing of the *T. maritima asd* (left) and *B. subtilis lysC* (right) lysine riboswitches. The overall pattern of chemical reactivity demonstrates the overall similarity in the global structure of two lysine riboswitch variants in their response to both magnesium and lysine. Red arrows indicate positions of ligand or Mg^{2+} dependent modulation.

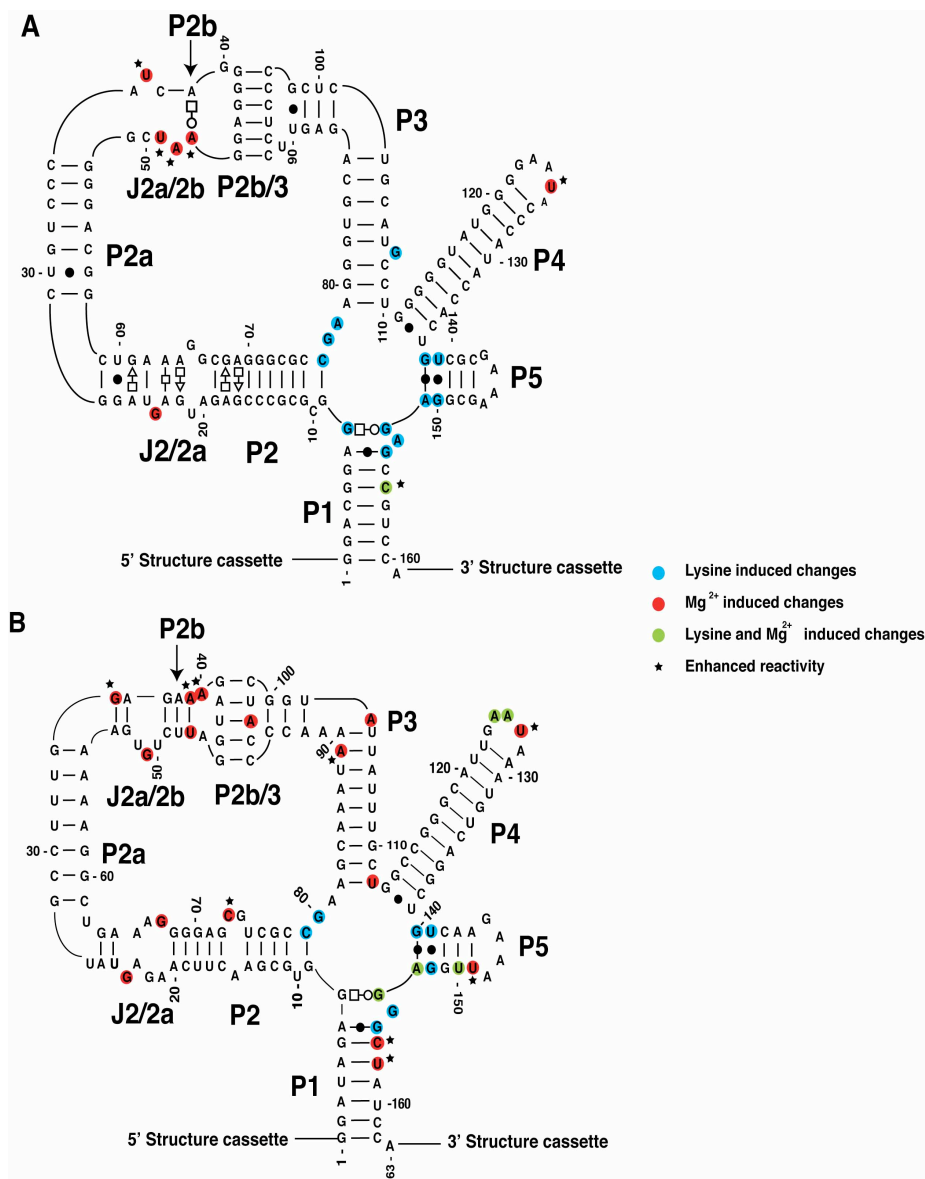


Figure 2.8 Sequence and structure of the RNAs used in SHAPE probing experiments.

The labeling of the paired regions is retained from figure 2.1. Non-canonical base pairing interactions shown in the structure of the *B. subtilis* variant are inferred from the crystal structure of the *T. maritima* variant. (A) Crystallographic construct of the *T. maritima* riboswitch used in the chemical probing studies shown in figure 2.7 (B) The riboswitch upstream of the *lysC* gene in *B. subtilis*. Note that the P5 region of this RNA was truncated to the same length as the *T. maritima* construct for consistency. Annotation of the reactivity patterns from Figure 2.7 are denoted on each RNA, per the legend to the right

general trends in the response of the lysine riboswitch to magnesium and ligand. First, both RNAs exhibit clear changes in reactivity upon addition of magnesium. These changes are centered on J2a/2b and tertiary interactions involving the terminal loops of P2, P3, and P4. For example, the conserved uridine in L4 (U125 *T. maritima* and U127 *B. subtilis* respectively) becomes highly reactive in both RNAs at 1-2 mM magnesium (Figure 2.7), reflecting its flipped out orientation upon formation of the P4 stem loop. This signature may also be indicative that the L4/P2 is formed at these Mg^{2+} concentrations. Second, both RNAs show limited lysine-dependent changes localized principally to the five-way junction--primarily J2/3, the base of P5, and 3'-side of the P1 helix (Figures 2.7 and 2.8). It is noteworthy that at high magnesium concentrations (>1 mM), lysine does not appear to significantly affect the structure in the *B. subtilis* variant which may reflect the truncations made to the P5 stem to mimic the *T. maritima* crystal construct (see methods). Lysine-dependent effects are however apparent for the *B. subtilis* RNA near physiological magnesium concentrations (0.2–0.5 mM). The protection patterns for both of these RNAs are in agreement with both in-line probing of the *B. subtilis* *lysC* riboswitch and an observed magnesium dependent fluorescence change of a 2-aminopurine label incorporated riboswitch at a position equivalent to A153 in the five-way junction¹⁰⁷. Together, these data indicate that the structure of the lysine riboswitch can be induced by magnesium to a near native state in the absence lysine at near physiological Mg^{2+} concentrations.

To further validate the chemical probing data, we analyzed the *B. subtilis* *lysC* riboswitch by small angle X-ray scattering (SAXS) in a collaboration with Rob Rambo at the Berkeley National Synchrotron light source. This solution method is sensitive to the

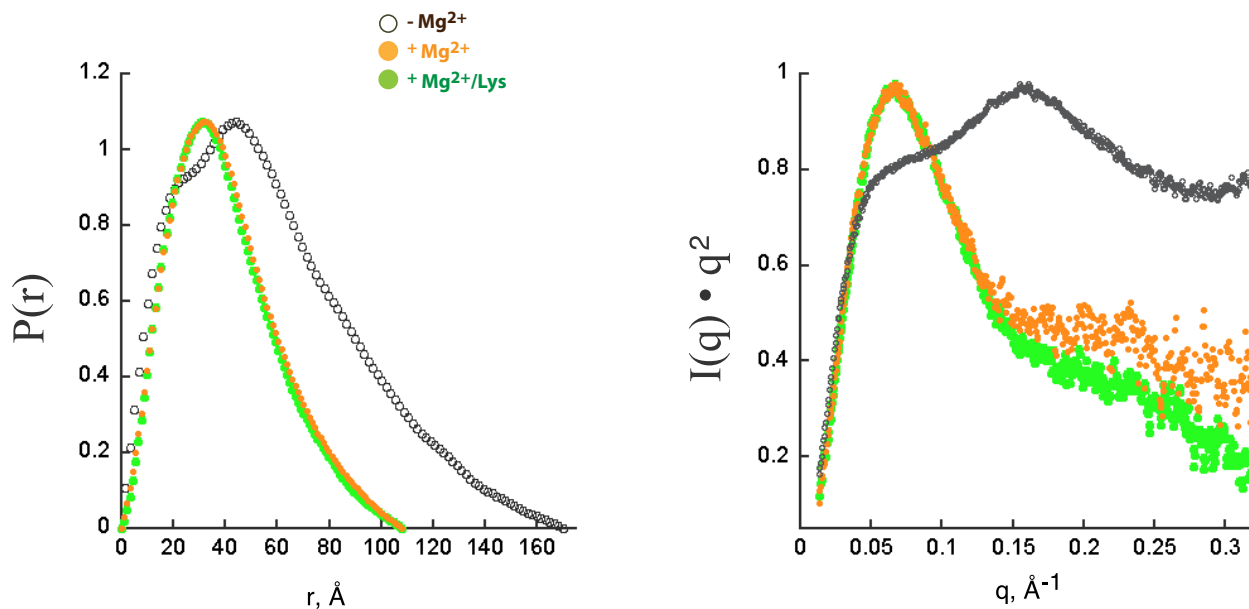


Figure 2.9 SAXS scattering curves of the lysine riboswitch

Small angle x-ray data corresponding to free (EDTA, gray; magnesium, orange) and lysine-bound (green). The left panel shows the experimental electron pair distribution plot where the x-intercept reflects the most likely maximum intermolecular scattering distance ($D_{\max}$). The right panel is a Kratky plot that reflects the extent of unfoldedness of the macromolecule {Russell:2002th}.

Table 2.2 Experimentally derived SAXS parameters

	Unfolded	Native	Native-Bound
parameters			
Guinier R_g , Å	41.8(±0.5)	29.7(±0.3)	31.1(±0.4)
Real space R_g , Å	47.3(±0.2)	32.0(±0.1)	31.5(±0.1)
$D_{\max}$ estimate, Å	172	108	108

Guinier R_g refers to radius of gyration determinations within the linear range of the scattering data when plotted as $\ln[I(q)]$ vs q^2 . Real Space R_g refers to radius of gyration determinations based on the integration of the pair distribution function for each data set.

global conformation of a macromolecule, and provides a direct measure of the radius of gyration (R_G), maximum dimension ($D_{\max}$), and electron-pair distribution function of the macromolecule¹³². In the absence of magnesium the riboswitch is weakly folded as illustrated by the hyperbolic feature of the Kratky plot (Figure 2.9), whereas the Kratky plot converges rapidly at small scattering angles reflecting a well-folded molecule in the presences of 5 mM Mg^{2+} ¹³³. This Mg^{2+} induced folding transition is also apparent from electron-pair distribution plots which describe the maximum dimension of the RNA, however the addition of lysine in the presence of 5 mM Mg^{2+} suggests that the dimensions of the free and bound conformations are identical (108-110 Å; Table 2.2). These dimensions are also consistent with the crystal structure after taking into account the first solvation shell (~108 Å). Together with the chemical probing data, these results indicate that this riboswitch can adopt a near-native form without lysine in solution with the aid of magnesium. Thus, lysine likely gains access to the binding pocket through small, localized fluctuations in the RNA structure as evidenced by chemical probing.

2.5- Temperature dependent folding of the lysine riboswitch

Another powerful feature of SHAPE chemistry is that the reactivity of the anhydride toward nucleophilic groups in the RNA can be monitored at various temperatures to allow more comprehensive sampling of the RNA folding trajectory than single temperature experiments⁷⁸. While the formation of ester adducts by NMIA with the RNA 2'-OH occurs more rapidly as a function of temperature, so too does quenching of the anhydride by the solvent, thus ensuring that the total reactivity is nearly identical over a wide range of temperatures⁷⁸. This approach can be leveraged to study the folding trajectories of an RNA in greater detail, as has been demonstrated by studies of

tRNA^{Asp78} and purine riboswitch folding⁸⁰. For the tRNA^{Asp} this analysis revealed that the secondary and tertiary structures of the RNA are strongly coupled as their unfolding occurs simultaneously, thus challenging the strict hierarchical view of RNA folding in which secondary structures are formed prior to tertiary structures⁸². Similarly, temperature dependent modification of the purine riboswitch showed that ligand binding imparts a large stabilizing effect on the three way junction as evidenced by a change of > 30°C in the T_m (the temperature at half the RNA becomes unstructured) for nucleotides in this region⁸⁰. In the context of particular mutations it was clearly demonstrated that ligand binding also promotes folding of the distal loop-loop interaction that stabilizes the global fold of the RNA, thus enabling these elements to exhibit highly cooperative folding.

Melting of the wild type lysine riboswitch from the *lysC* gene in *B. subtilis* was carried out in the presence or absence of lysine to determine how ligand binding reshapes the folding trajectory of this RNA. Probing was done under solution conditions similar to those already described with the exception that 50 mM potassium was included in these experiments to account for its participation in mediating lysine binding⁹⁵. In the absence of lysine the five-way junction regions were observed to be unstructured at all temperatures as indicated by the NMIA footprint (Figure 2.9). Addition of 1 mM Lysine results in strong protection of these junctions at temperatures below ~55°C, indicating that junction is the key site of ligand dependent changes as demonstrated by single temperature experiments. Just as observed for the guanine riboswitch⁸⁰, the five-way junction the lysine aptamer is by at least 30°C. Stability of the

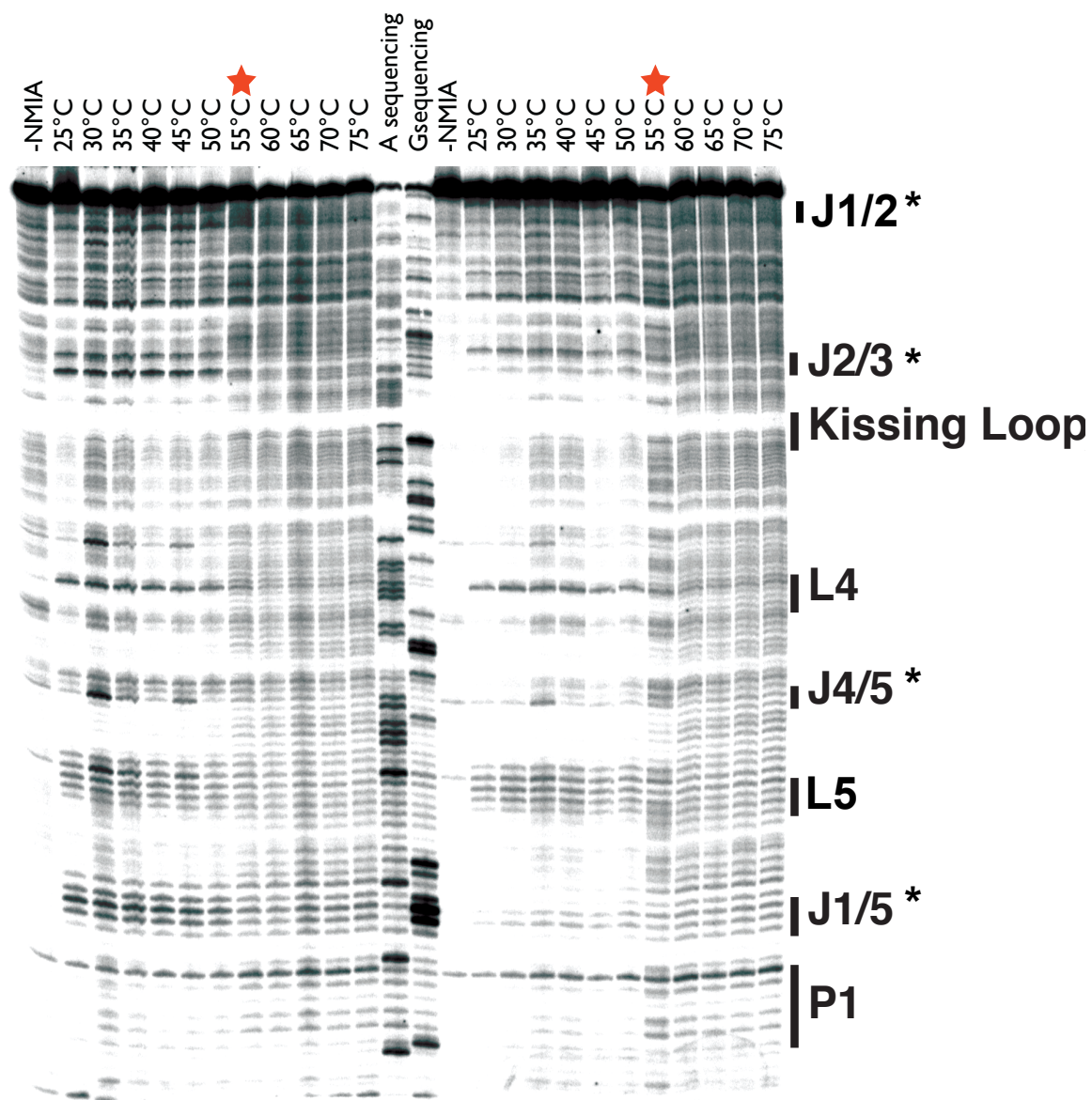


Figure 2.10 Temperature dependence of lysine riboswitch folding.

SHAPE chemical probing of the lysine riboswitch conducted from 25-75°C in the presence or absence of ligand. Asterisks indicate regions that are strongly protected by lysine binding. These regions coincide with the 5-way junction. In the absence of ligand this junction is disordered at all temperatures. Lysine binding induces the formation of an ordered structure up to ~55°C, a ΔT_m of > 30°C

global architecture of the RNA is however unaffected by ligand binding, arguing strongly that the aptamer adopts a stable scaffold over a range of physiologically relevant temperatures. As has been discussed in previous work the ability of the RNA to adopt preformed structures likely acts to restrict the conformational freedom of the ligand binding pocket to enable more rapid association kinetics. This view draws support from studies of the effect of Mg^{2+} on the ligand binding; Mg^{2+} primarily influences the observed rate of ligand association (k_{on}) rather than dissociation (k_{off})¹¹¹.

2.6- Resistance mutations in the lysine riboswitch

Resistance to the antimicrobial lysine analog AEC is conferred in both *E. coli* and *B. subtilis* via mutations within the lysine riboswitch (cyan, Figure 2.2A, Figure 2.11). These mutations fall into two general classes, mutations that directly abrogate contacts with lysine and those that appear to be important for folding of the aptamer. For example the universally conserved G8 and G9 nucleotide makes direct contacts with lysine that would be abrogated by mutations. Furthermore G8 participates in a conserved base quadruple interaction just below the ligand that may also be important for binding or regulation. More interestingly, others are observed in the distal regions of the P2 and P4 helix (nucleotides A62-A64 and G129), disrupting formation of key tertiary interactions. Notably, some of these mutants bind lysine with nearly the same affinity as the wild type *B. subtilis* lysC RNA⁸⁹, suggesting AEC resistance is gained by decreasing the rate at which the RNA folds into a binding-competent structure. As transcriptional regulation has a short temporal window in which to direct the

downstream secondary structural switch¹⁰⁹, a lowered folding rate would lead to loss of regulatory control. In this fashion, riboswitch-mediated regulation is governed by the rates at which individual elements are able to fold rather than interplay between the thermodynamic stabilities of each structure in the switch. Further studies will be needed to address the role of the ligand in the kinetic folding pathway of the riboswitch and its relation to efficient genetic regulation.

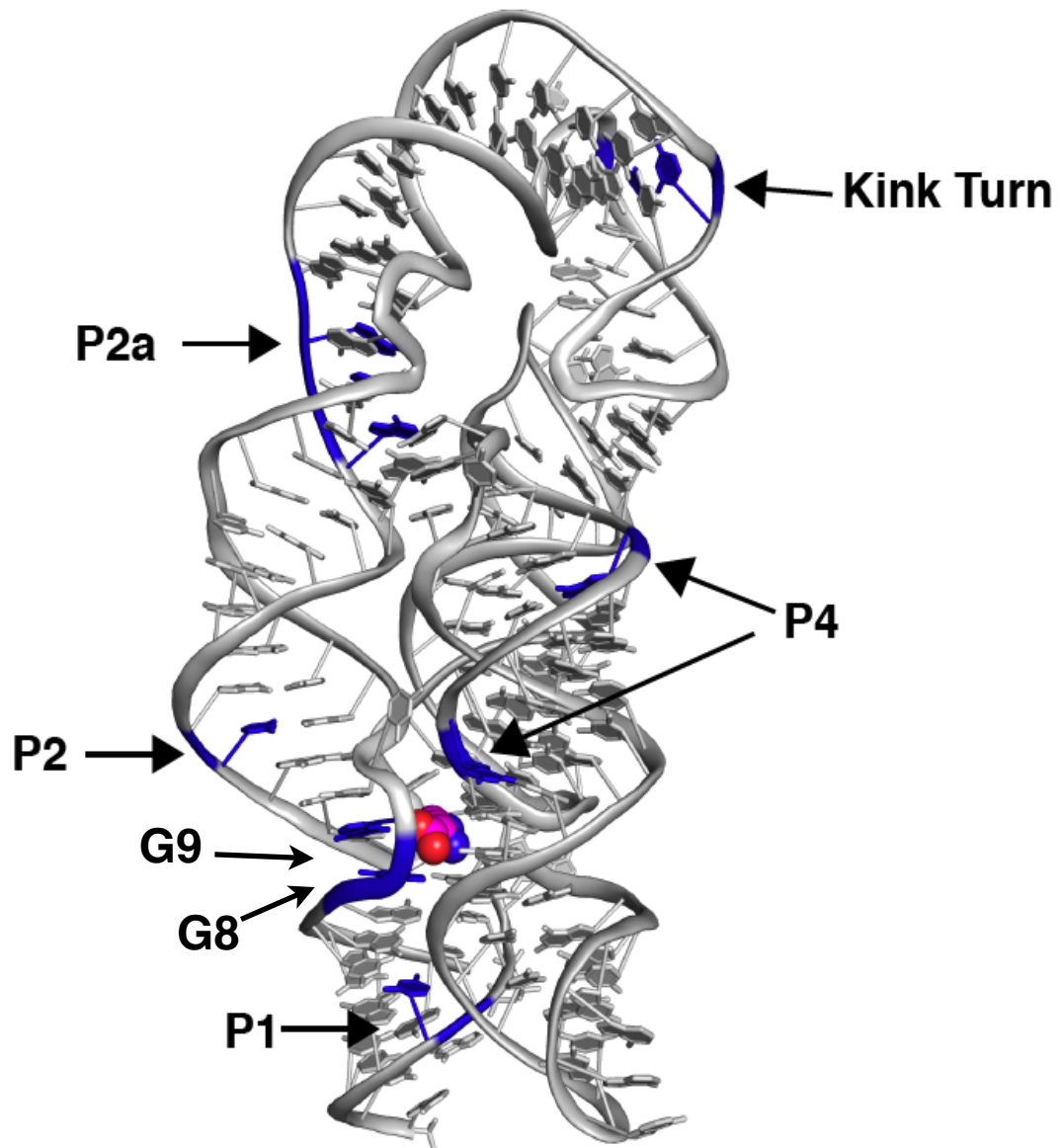


Figure 2.11 Mutations that confer AEC resistance

Mutations identified from AEC resistant *E. coli* and *B. subtilis* isolates that target the lysine riboswitch are highlighted in blue. G8 and G9 which make direct contacts to lysine are indicated.

2.7- Materials and Methods

RNA preparation and crystallization

A 161 nucleotide dsDNA coding for the *T. maritima* lysine riboswitch aptamer domain controlling the *asd* gene was constructed by PCR using overlapping oligonucleotides. The RNA was transcribed and purified using previously published techniques. The refolded RNA was then exchanged into 10 mM Na-HEPES, pH 7.0, 5 mM MgCl₂, and 2 mM lysine before storage at 4 °C. For the free state, RNA was refolded in the lysine supplemented buffer, exchanged three times into 10 mM Na-HEPES pH 7.0, 5 mM MgCl₂ followed by overnight dialysis into 1 L lysine-free buffer. The final concentration was determined by absorbance at 260 nm ($\epsilon=1,570,000 \text{ M}^{-1} \text{ cm}^{-1}$, MW=52,433 g mol⁻¹). RNA was stored at 4 °C until use.

The riboswitch was crystallized by the hanging drop vapor diffusion method in the presence of 1 mM lysine, or in the absence of lysine for the free state crystals. Drops were set up by mixing 1 μL of RNA with 1 μL of a mother liquor solution consisting of 2 M Li₂SO₄, 5 mM MgCl₂, and 10 mM Na-HEPES, pH 7.0, and 60 mM iridium hexamine to obtain the heavy atom derivative crystals. Identical conditions were used to grow the free state crystals, except no iridium hexamine was used in the mother liquor. Crystals were obtained within 24 hours and required no additional cryoprotection agent; they were looped with 2-3 μm loops and flash-frozen in liquid nitrogen before data collection.

X-ray data collection

Data for the bound state iridium hexamine derivative crystal was collected on beamline X29A at the Brookhaven National Synchrotron Light Source X-rays at the iridium absorption peak. This data was integrated, and scaled using HKL2000. All data used in phasing and refining came from a single crystal. Data for the unliganded structure was collected using CuK α wavelength (1.5418 Å) radiation on an R-Axis IV++ home source (Rigaku MSC), and the data was indexed and scaled using D*TREK.

Phasing and structure determination

Phases were determined by single wavelength anomalous diffraction (SAD) using data extending to 2.8 Å. SHELXD was used to find three iridium heavy atom sites within the asymmetric unit that had reasonably high occupancy. These heavy atom sites were used to calculate phases in SHELXE. The resulting experimental density map, following density modification (0.5 solvent fraction), displayed clear features corresponding to RNA backbone and base pairing (Fig S1). This map was used for initial building of the model.

The model was built in Coot and refined in PHENIX using iterative rounds of building and refinement. The RNA nucleotides were initially built along with four iridium hexamine molecules. This model was brought through multiple rounds of simulated annealing and atomic displacement factor (ADP) refinement before building lysine into the model. At this point, the density for the entire ligand was clearly visible and was validated by inspection using a simulated annealing omit map in which the ligand and a few surrounding nucleotides were omitted from the model (Figure S2). One round of water-picking was carried out by the PHENIX ordered solvent protocol; waters were

chosen based on peak size in an F_o-F_c map. R_{free} was monitored in each round to ensure that it was dropping. Figures were prepared using and Pymol.

The unliganded RNA model was built using the bound form as a molecular replacement solution using only the RNA and re-refined using iterative rounds of simulated annealing and ADP refinement. The final refinement statistics are shown in Table S1, and the structure factors and models have been deposited in the Protein Data Bank (accession codes 3D0U and 3D0X).

Chemical probing using Selective 2'-Hydroxyl Acylation analyzed by Primer Extension (SHAPE) chemistry

RNA sequences were constructed to correspond to the *T. maritima* RNA that was crystallized as well as the riboswitch controlling the *B. subtilis* *lysC* gene. The *B. subtilis* sequence was truncated in the P5 region to match the length of the *T. maritima* sequence to ensure that the RNAs are comparable, and the 5' and 3' structure cassettes were appended to these sequences as described previously. RNA was generated by run off transcription and purified according to the same protocol used to generate RNA for the crystallographic studies.

RNA was prepared for modification by placing 1 μL of 2 μM RNA (2 pmols) into 11 μL 0.5x T.E. buffer. This sample was heat/cooled to allow the RNA to refold and then supplemented with 6 μL buffer consisting of 333 mM K-HEPES, pH 8.0, and 333 mM NaCl. This buffer was supplemented with 2 mM lysine for the plus ligand reactions (final concentration of 667 μM), and MgCl_2 was included in the folding buffer at concentrations ranging from 6.8 mM to 425 μM in 2 fold dilutions to yield the concentrations shown in the magnesium titration experiments. All reactions were supplemented with 1 μL 3 mM Mg^{2+} to ensure proper extension by the polymerase during the reverse transcription step. Modifications were carried out using 30 mM NMIA for the recommended five half lives at 20 $^\circ\text{C}$; reverse transcription and gel running procedures were preformed as previously described.

Small angle X-ray scattering (SAXS) data collection and analysis

The *B. subtilis* *lysC* riboswitch was prepared for SAXS analysis using an Ettan LC liquid chromatography system configured with a Superose 6 PC 3.2 size-exclusion column (GE Healthcare), in-line vacuum degasser and a 0.01 μm filter. Three distinct sample buffer conditions were used and are referred to as native-bound, bound and unfolded containing either 2 mM lysine and 5 mM MgCl_2 or 2 mM EDTA, respectively. All sample buffer conditions contained 20 mM Na-HEPES, pH 6.5 and 50 mM KCl. Prior to gel-filtration purification of the sample the column was equilibrated with the appropriate buffer. Likewise, RNAs were refolded as described above at 5 μM and concentrated to 50 μL to a final concentration of 5 mg/mL. Each purification utilized 50 μL of sample and a fraction corresponding to the major eluting peak was taken for direct SAXS experiments.

SAXS data was collected at the synchrotron beamline 12.3.1 of the Advanced Light Source (Berkeley, CA). All scattering data was collected at 1.0332 $\text{\AA}$ using a minimal sample volume of 18 μL . A full scattering curve was measured as two separate exposures at 6 and 60 s for the sample and buffer. The X-ray scattering due to the riboswitch RNA was determined by subtracting the background X-ray scattering of the

gel-filtration buffer from that containing the RNA and buffer. Data was reduced and combined with Primus to produce the final X-ray scattering curve. Data was collected over a range of RNA concentrations (0.7 to 0.1 mg/mL). No concentration dependent changes were observed in the lowest scattering angles.

The radius of gyration (R_g) which describes the mass distribution of a particle around its rotational center of mass was determined using either the Guinier approximation (Guinier R_g) within the angular range (q) of $q \cdot R_g < 1.3$ or by calculating the electron pair-distribution function (Real Space R_g) in Gnom. All final plots were prepared with KaleidaGraph.

Chapter 3: Correlative examination of the thermodynamics, kinetics and regulatory response of the lysine riboswitch

Considerable effort has been invested in deciphering the structural features of riboswitch aptamer domains that enable the mRNA to bind the effector with both high affinity and specificity to achieve precise genetic control against a background of highly similar chemicals in the cell. For example, purine riboswitches distinguish between guanine and adenine with >20,000 fold specificity based on the identity of a single pyrimidine that base pairs with the purine ligand^{60,92}. Likewise, the various classes of S-adenosylmethionine (SAM) binding riboswitches discriminate SAM from S-adenosylhomocysteine (SAH), a toxic byproduct of SAM methylation reactions, through electrostatic interactions with the sulfonium ion of SAM^{69,134,135}. Displacement of critical carbonyls away from the sulfonium ion leads to a dramatic loss in the ability of the SAM-I riboswitch to discriminate between the two compounds¹³⁴. Specificity can also be mediated in part by associated solvent and ions tightly coordinated with the ligand-RNA interaction, as observed in the FMN and lysine riboswitches^{93,95}. As many riboswitches control the production or uptake of metabolites critical for normal growth and reproduction these RNAs are considered to be potentially important therapeutic targets⁴⁶. Thus, a deep understanding of how specific recognition is achieved by riboswitches is essential to guide rational design of novel compounds that target these RNAs.

To act as an effective therapeutic, a compound must not only productively bind to the aptamer but also exert control over the regulatory switch. A recent study of a tetrahydrofolate (THF) riboswitch revealed that while THF and guanine bind its aptamer

domain with nearly identical affinities (K_D), only the folate is capable of terminating transcription. For riboswitches that act on transcription that only have a limited timeframe (Δt_{RNAP}) to sense its environment before the RNA polymerase (RNAP) proceeds beyond the intrinsic terminator, at which point the riboswitch loses regulatory control^{136,137}. This results in a kinetically constrained process in which the effector concentration required to achieve half maximal transcriptional regulation (T_{50}) may not correlate to binding affinity due to insufficient time for ligand binding to reach equilibrium. For example, a detailed analysis of the flavin mononucleotide (FMN) riboswitch demonstrated that the T_{50} for this RNA is ~10-100 fold in excess of the K_D , which was attributed to the slow binding kinetics of FMN¹⁰⁹. While it is tempting to generalize this observation to other riboswitch classes¹³⁸, other factors such as programmed pausing¹³⁹, the presence of protein factors NusA and NusG^{140,141}, and the intracellular NTP concentration can influence this process^{109,142}. Thus, to fully assess how the natural effector and other compounds act upon the riboswitch, the relationship between binding kinetics and regulatory activity must be established.

To fully explore the relationship between equilibrium affinity (K_D), binding kinetics (k_{on} and k_{off}) and regulatory activity (T_{50}), I have examined the mechanistic details of the *Bacillus subtilis* *lysC* lysine-responsive riboswitch. The aptamer of this riboswitch folds into a complex structure involving the packing of multiple helices together that organize a binding pocket for lysine within a five-way junction (Figure 2.2 and Figure 2.4)^{95,143}. The K_D and binding kinetics for lysine and a number of related analogs were examined that demonstrate how this RNA achieves lysine specificity against a background of abundant amino acids found that bear similar main chain functionality. These data

provide a benchmark for comparative analysis of the aptamer binding affinity and kinetics to transcriptional regulation, allowing us to examine the relationships between these parameters. Interestingly, while previous studies suggested that the T_{50} of lysine is ~1000 fold greater than the K_D ^{89,107,107,144} I show that lysine and related compounds promote efficient transcriptional termination within ~3-fold near their K_D under NTP concentrations similar to those previously reported. This finding is consistent with simplified kinetic models of transcriptional regulation based on the observed binding rate constants. Importantly, all measurements were conducted using closely matched buffer conditions, in part motivated by structural and biochemical data that demonstrated the importance of K^+ and Mg^{2+} for lysine binding⁹⁵, which was not the case in previous studies of this riboswitch^{22,89,107,144}. However, at NTP concentrations of $\geq 400 \mu M$ regulation requires a ~5 fold increase in lysine suggesting that the regulatory efficiency of this riboswitch could be indirectly tuned by the metabolic state of the cell, such as under nutrient limiting conditions. Under high NTPs the riboswitch also exhibits a greater degree of selectivity for lysine compared with the majority of analogs, with the exception of two of the main chain derivative whose rapid binding kinetics render them relatively insensitive to the transcription speed.

3.1- Characterization and validation of a 2-aminopurine labeled receptor.

To establish the apparent equilibrium dissociation constant ($K_{D,app}$) for each of the compounds used in this study, I site-specifically labeled a two-piece aptamer derived from the *B. subtilis* *lysC* riboswitch. A fluorescent 2-aminopurine (2AP) nucleotide was incorporated at a non-conserved position in the five-way junction (G164, Figure 3.1A)

that chemical footprinting demonstrated undergoes a significant conformational change upon lysine binding^{22,143} and is solvent exposed in the crystal structure^{95,143}. These data imply that this nucleotide likely transitions from being stacked with other nucleotides in the unbound aptamer to becoming solvent exposed upon lysine binding. This was suggested by a previous study from the Lafontaine group that also used a 2AP reporter at this position to monitor the effects of Mg^{2+} on global folding of the RNA¹⁰⁷. As expected, a strong increase in fluorescence was observed upon the addition of lysine (Figure 3.1B), consistent with the labeled position becoming flipped out into solution upon binding. Fitting the data to a simple two-state binding model yielded a $K_{D,app}$ for lysine of $53 \pm 5 \mu M$ at 37 °C.

The affinity of lysine for the 2AP reporter is significantly weaker than previous reports^{22,89,95,105}. However, folding and ligand binding by the lysine aptamer have been shown to exhibit a strong dependence upon the concentration of Mg^{2+} ^{95,107,145}. To determine whether the decline in affinity was due to changes in the RNA or differences in the experimental conditions I measured the affinity of lysine for the wild type *B. subtilis* *lysC* aptamer domain (nucleotides 30-199 from transcription start site) under the same experimental conditions using isothermal titration calorimetry (ITC). A $K_{D,app}$ of $18 \pm 2.2 \mu M$ was measured for lysine at 37 °C (Figure 3.1C), within ~3-fold of the 2AP reporter, indicating that the lower affinities observed here are primarily due to the experimental conditions chosen for these studies. The measured $\Delta H^\circ = -21 \pm 1.6 \text{ kcal}\cdot\text{mol}^{-1}$ and $-\Delta S^\circ = 15 \pm 2.0 \text{ kcal}\cdot\text{mol}^{-1}$ are comparable to values obtained of other riboswitch aptamer-ligand interactions^{60,105,134} and agree well with previous ITC data of the lysine riboswitch derived from *Clostridium acetobutylicum*¹⁰⁵, reflecting a highly

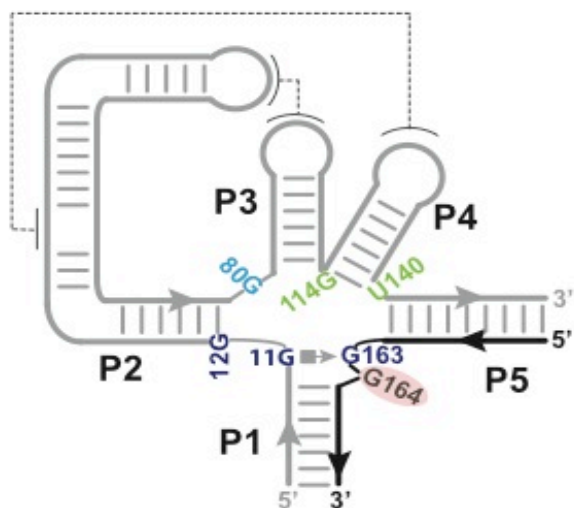
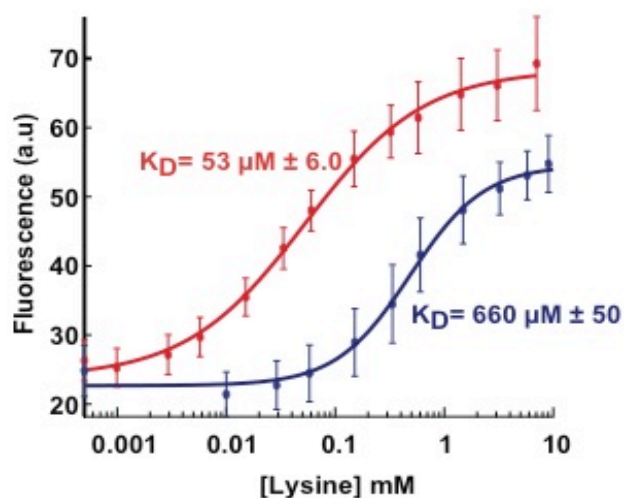
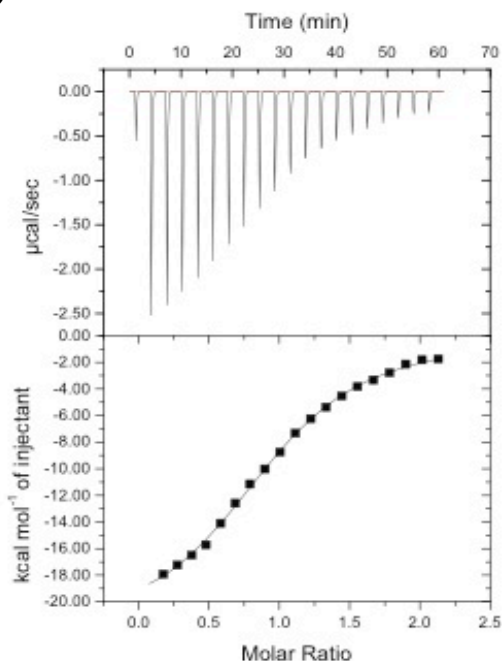
A**B****C**

Figure 3.1 Validation of the 2AP reporter for determination of ligand binding affinities.

A) Illustration of the two-piece 2AP reporter construct utilized for the binding studies presented in this chapter. The gray and black strands were annealed as described in methods to reconstitute a functional RNA for fluorescence based assays. B) Sample fluorescence data from lysine titrations with the 2AP labeled construct in the presence (red) or absence (blue) of 10 mM KCl. Note the decrease in both the binding affinity and change in fluorescence upon omission of KCl from the buffer conditions. C) Lysine binding to the wild type lysC aptamer as measured by ITC demonstrates reasonable agreement between the K_D values for the 2AP labeled construct and the native RNA

enthalpic binding event.

As second test of the validity of the 2AP reporter the effect of potassium on lysine binding was measured. This ion was found to mediate recognition of the α -carboxyl group of lysine (Figure 3.2)⁹⁵. Under these conditions, substitution of potassium for sodium results in a ~ 10 -fold decrease in affinity (Figure 3.1B) confirming the importance of this monovalent cation in mediating a high affinity interaction. Interestingly, comparison of absolute fluorescence between the two monovalent cations suggests that the K^+ selectively stabilizes the flipped out conformation of the 2AP reporter as evidenced by the increased change in fluorescence at saturating lysine (Figure 3.1B).

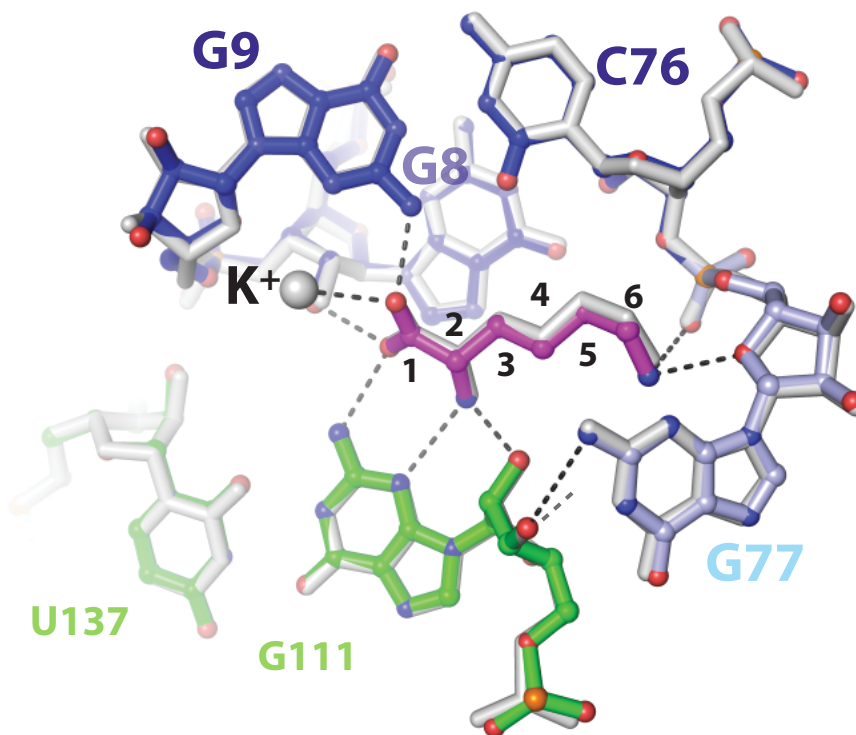


Figure 3.2 Important contacts of the lysine riboswitch binding pocket.

Overlay of the lysine riboswitch structures from Garst *et. al.* 2008 (colored, PDB ID: 3D0U) and Serganov *et. al.* 2008. (gray, PDB ID: 3DIL). Note the latter structure was solved in solution conditions that allowed for identification of a key potassium ion. Waters are omitted for simplicity.

This interpretation consistent with primer extension assays that demonstrated a potassium-dependent stop that occurs only in the presence of lysine⁹⁵, and suggests that this ion is recruited specifically to the RNA-ligand complex as opposed to stabilizing preformed structures in the unliganded aptamer.

3.2- Molecular determinants of specificity

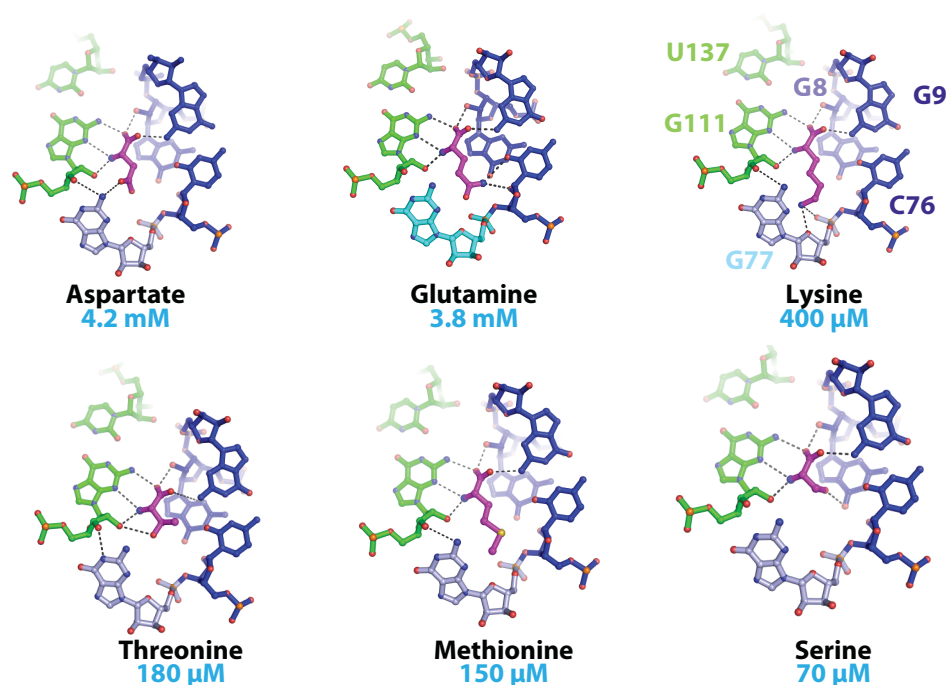
In their cellular context, riboswitches must contend with a complex chemical environment that challenges their ability to coordinate gene expression in response to a specific effector. Recent studies have provided evidence that some riboswitch aptamers can bind alternative cellular ligands. For example, binding different hexose sugars have been demonstrated to inhibit glucose-6-phosphate dependent cleavage by the *glmS* riboswitch *in vitro* and *in vivo* enabling the riboswitch to mediate chemical crosstalk between related compounds¹⁴⁶. While selectivity of the lysine riboswitch appears to be determined by an electrostatic interaction with the ϵ -amino group of lysine (Figure 2.4 and Figure 3.2), other amino acids could mediate identical main chain interactions with the RNA. As amino acids are the most highly concentrated metabolites in the cell¹⁴⁷, suggesting the possibility of promiscuous binding. Docking of the other amino acids into the binding pocket, which is assumed to be rigid, indeed suggests that some side chains that could potentially form favorable interactions with the RNA (Figure 3.3).

To determine whether the lysine aptamer can productively bind other amino acids, I tested a subset of standard amino acids that have been demonstrated to have high intracellular concentrations¹⁴⁷, and can be readily docked within the binding site of the RNA. Most of these amino acids, including alanine, asparagine, threonine and

methionine, showed no fluorescence change in the 2AP labeled aptamer (Figure 3.4). Only serine displayed some binding, albeit with extremely low affinity (>20 mM), well above its intracellular concentration. These results are consistent with previous reports that ornithine and norleucine, non-standard amino acids that have a shorter side chain by a single carbon unit or lack the primary amine, respectively, are unable to bind the lysine riboswitch aptamer^{22,89}. Thus, the backbone atoms, while certainly contributing to binding affinity of lysine, are not sufficient to confer binding independent of the side chain functionality. Positioning of the amino group adjacent to the G77 phosphate and O4' ribosyl oxygen must be essential for binding, providing the basis for discrimination against the abundant amino acids present in the cell.

A set of compounds that alter key aspects of the lysine functional groups was next employed to further investigate the energetic contributions of particular features of the binding site (Figure 3.4). From previous crystallographic studies there are two sets of interactions between the ϵ -amino group and the RNA hypothesized to be important: an electrostatic interaction between the amine and the non-bridging phosphate oxygen of G80 and a water-mediated interaction between the amine and the 2'-OH of G111. Since the 2'-OH of G111 also interacts with the α -amino group of lysine, the hydroxyl group acts a hydrogen bond donor to this water (W1, Figure 2.1). The importance of the water was tested using *N-methyl lysine* (NML), which maintains the position and charge of the side chain amine, but would have to displace W1 in order to accommodate the additional methyl group in a rigid pocket. This compound displays a moderate loss (~34-fold) in affinity relative to lysine, verifying that this water is energetically important for ligand binding.

A



B

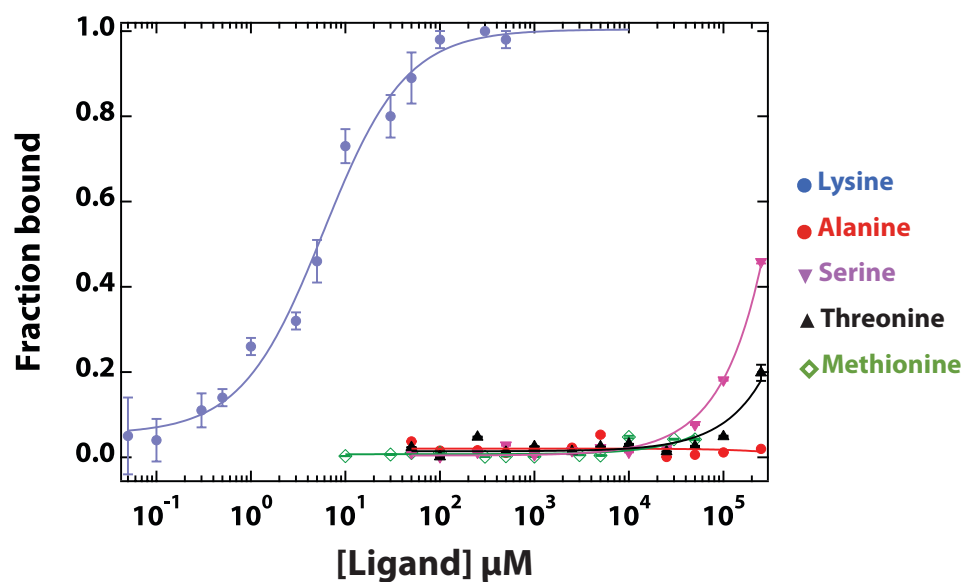


Figure 3.3 Test of standard amino acids for binding competency

(A) Modeling of alternative amino acids in the lysine binding site. The models are arranged in order cellular abundance of the amino acid in *E. coli* during exponential phase growth 35 as indicated by the blue labeled values. Amino acids were modeled by replacing the lysine side chain in the Pymol (Delano Scientific) and sampling side chain rotamers for complementarity. (B) Normalized fluorescence data showing lysine in blue and alternative amino acids as labeled. As indicated in the text, I do not observe appreciable binding under the conditions tested.

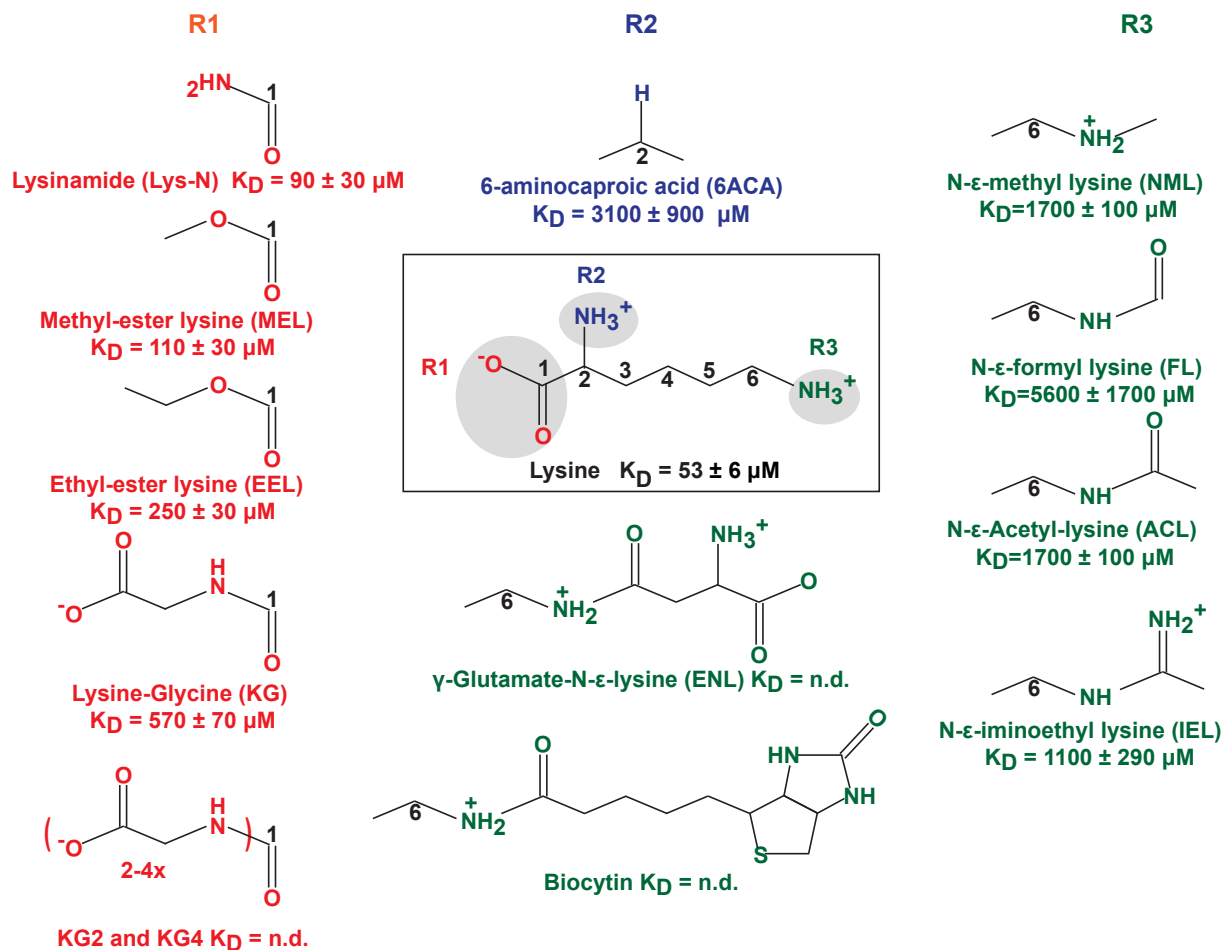


Figure 3.4 Energetics of molecular recognition.

The functional groups that were directly tested in this study are labeled R1, R2 and R3 corresponding to the carboxylate, main chain amino and ε-amino groups respectively. The brackets next to the full name of each compound indicate the abbreviations used to refer to each ligand in the remaining text. Additions and substitutions of R1 have modest effects on the binding affinity indicating that this functional group has weak energetic contributions to binding. Deletion of R2 or the main chain amine carries a more pronounced energetic penalty, but is not an absolute prerequisite for binding. Modifications that alter the charge character of the ε-amino or disrupt water mediated interactions influence binding to a similar degree.

The importance of the presence and position of a positive charge was tested using a set of three compounds (Figure 3.4, right). *N*- ϵ -formyl lysine (FL) and *N*- ϵ -acetyl lysine ablate the positive charge on the nitrogen and add steric bulk to the *N*- ϵ group. Structures of homoarginine and *N*- ϵ -iminoethyl lysine (IEL) bound to this RNA suggest that the additional functionality of these compounds can be accommodated in the pocket without any steric clash and with the carbonyl oxygen positioned in place of W1 such that it accepts a hydrogen bond from the 2'-OH of G111. Surprisingly, despite lacking a formal charge on the side chain, these compounds exhibit only a 4.5-fold loss in affinity relative to the positively charged compound NML. This clearly shows that the positive charge of the amino group is not essential for productive binding. Most of the energetic penalty for these compounds is due to their additional steric bulk; IEL, which can localize positive charge to the ϵ -amino group binds only ~2-fold better than NML and ~10-fold worse than lysine. Together, these data argue that both positive charge and water-mediated interactions are important to obtain the full binding affinity of lysine. Removal of both, as in norleucine, completely ablates binding. Also, addition of very large functional groups to the ϵ -amino group of lysine, such as those found in biocytin and γ -glutamate- ϵ -*N*-lysine are not tolerated, indicating that the binding pocket cannot rearrange to accommodate bulkier modifications to lysine.

The main chain functional groups of lysine, and in particular the α -carboxyl group, appear to play a lesser role in establishing a high affinity interaction with the lysine riboswitch aptamer domain. 6-aminocaproic acid (6-ACA) binds the RNA, albeit with weak affinity, despite lacking the main chain amine which H-bonds with the N3 and 2'-OH groups of G114 (Figure 3.2). Even more strikingly, modifications to the α -carboxyl

group are well tolerated. Conversion of this functional group to a neutral amide in lysinamide causes no loss in binding affinity, despite its likely inability to recruit the K^+ ion to the site adjacent G9 (Figure 3.2). Likewise, a methyl ester at this position only drops binding affinity by ~3-fold. Further additions to this group with sterically bulky moieties (ethyl- and glycyl-) diminish affinity, but not nearly to the same extent as modifications to either of the two amino groups. However, beyond a single glycine (KG dipeptide) the RNA cannot accommodate larger peptides (KG₂ – KG₄), suggesting that these modifications would introduce steric clashes that cannot be tolerated by the RNA.

3.3- Evidence for rigidity of the lysine binding site

To determine whether binding different analogs has any effect upon the global or local structure of the RNA I employed a chemical probing technique known as “SHAPE”⁸¹. This method generally assesses local flexibility of the backbone with single nucleotide resolution based on reactivity of ribose 2'-hydroxyl groups with N-methylisatoic anhydride (NMIA). Reactivity of NMIA is determined by the conformation of the ribose sugar bearing the 2'-hydroxyl group with the C2'-endo configuration being the reactive state and has been demonstrated to strongly correlate with the NMR order parameter¹⁴⁸. Nucleotides in the five-way junction exhibit high reactivity towards NMIA in the absence of ligand indicating that this region of the RNA is conformationally flexible. Addition of saturating ligand concentrations (1 mM lysine or 10 mM of the other compounds) I observe the same set of chemical reactivity protections for each of the junction segments (Figure 3.5) as well as the chemical reactivity pattern globally that is consistent with formation of the crystallographically observed tertiary architecture.

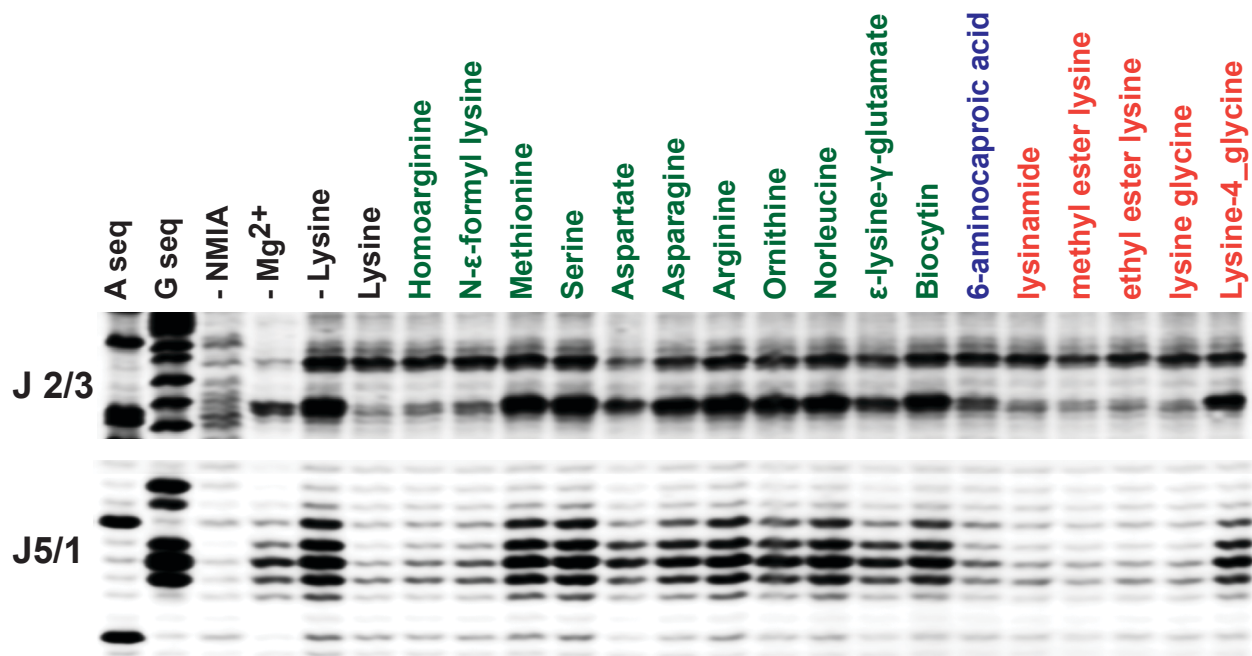


Figure 3.5 Chemical probing of the lysine riboswitch

Chemical modification of the lysine riboswitch in the presence of either 1 mM lysine or 10 mM of a subset of the analogs from this study. In the absence of ligand (-Lysine lane) the high reactivity of the J2/3 and J5/1 regions denote local flexibility of the RNA backbone in this region. Addition of binding competent ligands induces conformational changes that stabilize the binding pocket and constrain these nucleotides in non-reactive conformations, as indicated by the observed protection patterns. The similarity of the protection patterns of induced by binding competent ligands indicates that the RNA adopts a conformation similar to that observed in the lysine-RNA complex. The differences in the relative amount of protection reflect a difference in the fraction bound at the ligand concentrations used for this experiment. Labels colored according to the R1, R2, and R3 scheme in figure 3.4.

These data strongly argue that the RNA structure adopts a similar conformational state upon binding structurally distinct compounds. This is in agreement with structures of side chain analogs such as homoarginine and IEL that show an identical conformation of the RNA despite additions to the ϵ -amine⁹⁵.

While previous structural studies of the lysine riboswitch have focused on side chain derivatives of lysine, the ability of many main chain compounds to bind with reasonably high affinity raises questions about how this apparently rigid binding pocket can accommodate these different chemical alterations. To address these questions I determined the crystal structure of the 6-ACA and KG compounds in complex with the *T. maritima* aptamer using X-ray crystallography. These crystals diffracted to ~ 3.0 Å resolution and models were generated by molecular placement using the free state lysine riboswitch structure (PDB ID: 3D0X). Maximum likelihood superimposition of the refined models with the original lysine bound RNA using THESEUS¹⁴⁹, revealed the RNA adopts an identical fold when bound to these alternative ligands as predicted by SHAPE (Figure 3.5). The average all-atom displacement of residues within the binding pocket was less than the coordinate error of the structure.

These structures revealed that these two compounds oriented to maximize hydrogen bonding interactions with the RNA. In the structure of the KG-RNA complex, the glycine residue of the dipeptide is clearly identifiable in Fo-Fc difference maps (Figure 3.6), extending an additional 3.6 Å away from the central cavity of the RNA. The terminal carboxyl group of the dipeptide hydrogen bonds to the 2'-hydroxyl group of G112 at the base of the P4 helix. Interestingly the interaction replaces an ordered solvent molecule that coordinates to the potassium observed in the lysine bound

Supplemental Table 1

Ligand in complex	Lysine-glycine (KG)	6-aminocaproic acid (6ACA)
Data Collection		
Space group	P32	P32
Cell dimensions		
a, b, c (Å)	119.9,119.9,58.7	119.9,119.9,58.7
α, β, γ (°)	90,90,120	90,90,120
Wavelength (Å)	1.5418	1.5418
Resolution (Å) ^a	19.62-3.00(3.10-3.00)	59.8-3.0(3.1-3.0)
Rsym or Rmerge (%) ^a	0.113(0.321)	0.098(0.338)
$I/\sigma I$ ^a	5.1(1.9)	5.4(1.7)
Completeness (%) ^a	97.8(96.5)	97.7(99.9)
Redundancy ^a	1.92(1.93)	1.59 (1.58)
Refinement		
Resolution (Å) ^a	19.6-3.0(3.2-3.0)	26.7-3.0(3.1-3.0)
No. Reflections	18033	17611
Rwork / Rfree	18.4/23.7	18.3/22.3
No. Atoms	3503	3490
RNA	3479	3479
Ligand	14	9
Water	10	2
Atomic Displacement Parameters		
All	48.1	55.5
RNA	48.1	46.7
Ligand/ion	53.8	36.7
Water	31.8	50
R.m.s. Deviations		
Bond lengths (Å)	0.018	0.006
Bond angles (°)	1.68	0.914
Maximum likelihood coordinate error (Å)	0.38	0.34

a Values for the highest resolution shell are shown in parentheses.

b $R_{\text{merge}} = \sum |I - \langle I \rangle| / \sum I$.

c Refinement was against all data within the stated resolution range, with a random 10% omitted for use in calculation of Rfree.

d $R_{\text{xtal}} = \sum ||F_o| - |F_c|| / \sum |F_o|$ for the working set. Rfree is the same for the test set.

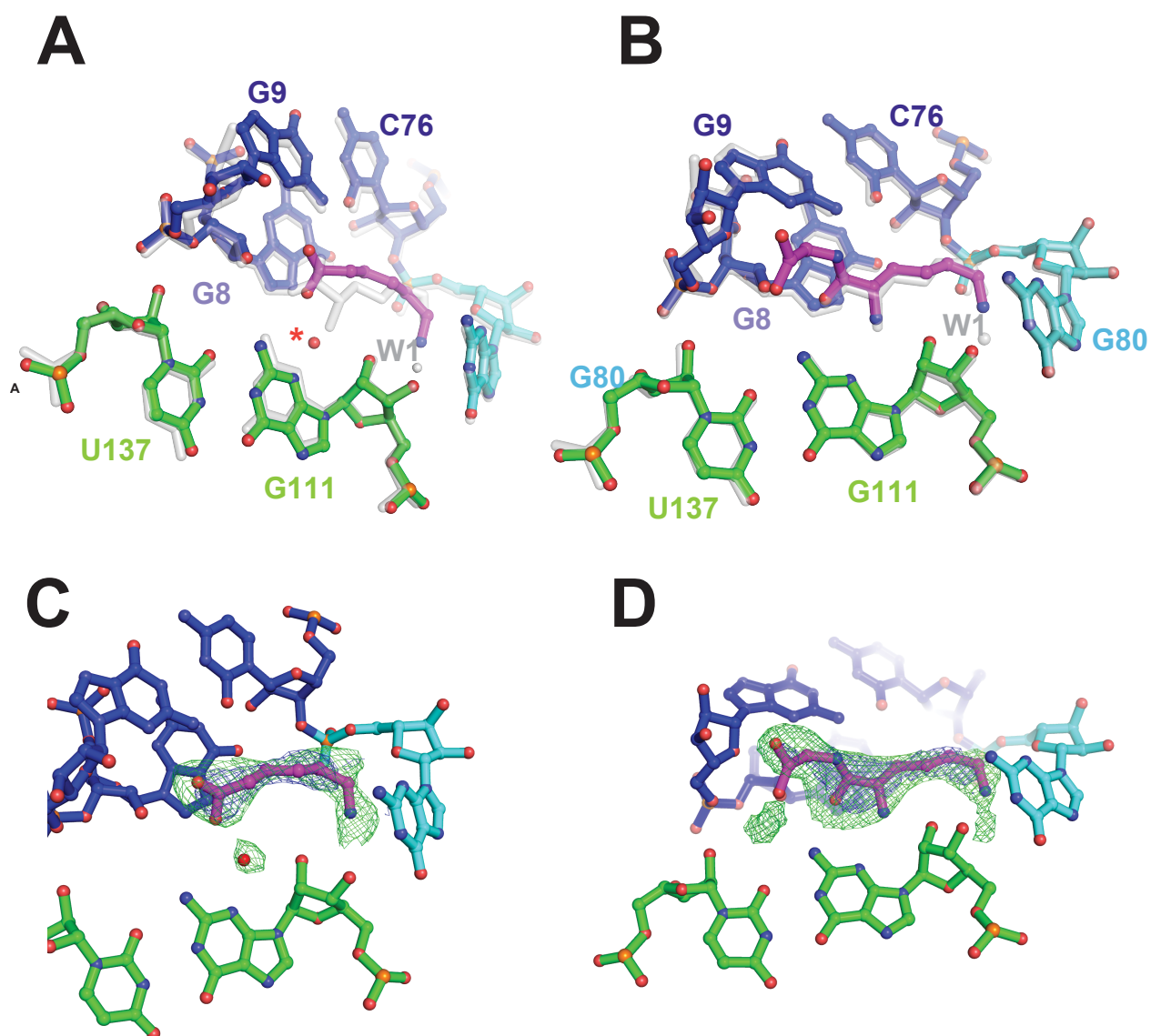


Figure 3.6 Lysine binding pocket enforces strict maintenance of H-bonding interactions
 The structure of A) 6ACA and (B) KG compounds bound to the *T. maritima* lysine riboswitch. Structures are overlaid with the lysine bound structure from Serganov *et al.* (light gray) for reference. The RNA was found to adopt an identical conformation upon binding these distinct ligands, as implied by chemical probing. The simulated annealing omit maps are shown in (C) and (D) for reference.

structure indicating that water mediated contacts are important in multiple facets of the binding interface. In the 6-ACA bound structure, two distinct peaks of density were observed in the Fo-Fc maps between the ligand and G111 (Figure 3.6). These peaks were modeled as water and are near the position occupied by the α -amino group of lysine. The carboxyl group of 6-ACA is rotated slightly to hydrogen bond with this water that also forms hydrogen bonding interactions with the sugar edge of G111 at the base of P4. These structures provide no evidence for structural adaptation in the binding pocket, but rather indicate that the constraints of this site force adaptation by the ligand to the binding pocket through new hydrogen bonding interactions or recruiting ordered solvent to mediate ligand-RNA contacts. This is consistent with previous results from the Serganov *et. al.*⁹⁵ in which the various side chain analogs stabilized an identical RNA conformation, and likely reflects the stringency of the structural constraints placed on the ligand binding site of this RNA. In contrast, nucleotides within the three way junction of the purine riboswitch were found to undergo local conformation shifts to accommodate alternative ligands¹⁵⁰. The degree to which different aptamers exhibit this structural plasticity is therefore likely to be idiosyncratic amongst various riboswitch classes.

3.4- Binding kinetics suggest a rapid approach to equilibrium.

The activity of riboswitches that regulate transcription have been shown to be influenced by the kinetic properties of ligand binding such that the concentration of effector required to elicit the regulatory response (T_{50}) can be significantly higher than the concentration needed to half-saturate the aptamer domain (K_D)¹⁵¹. To define the

relationship between binding and activity of the *B. subtilis* *lysC* riboswitch the kinetic parameters of lysine binding were measured using stopped flow experiments that monitor changes in fluorescence of the 2AP labeled lysine aptamer as a function of time. Lysine association and dissociation was measured under identical conditions as K_D measurements (buffer; 37 °C) using pseudo-first order conditions in which the ligand is in excess over the labeled RNA¹⁵². The observed rate of association (k_{obs}) as a function of lysine concentration yield an estimate of $k_{on} = 4600 \pm 200 \text{ M}^{-1}\text{s}^{-1}$ and $k_{off} = 0.2 \pm 0.06 \text{ s}^{-1}$ (Figure 3.7A). Direct measurement of k_{off} was also determined by monitoring fluorescence decrease upon rapid dilution of the bound complex yielding a comparable value ($0.4 \pm 0.01 \text{ s}^{-1}$ (Figure 3.7B).

A previous study¹⁰⁹ demonstrated that a simple model of ligand binding kinetically controlled regulation; a theoretical prediction of the T_{50} response can be related to the observed rate of binding (k_{obs}) using the following equation¹¹².

$$T_{50} = \frac{[RL]}{[R]_T} (\Delta t_{RNAP}) = \left(\frac{[L]}{[L] + K_D} \right) * (1 - e^{-k_{obs} \Delta t_{RNAP}})$$

where $([RL]/[R]_T)(\Delta t_{RNAP})$ is the fractional saturation of the aptamer at time Δt_{RNAP} and $[L]$ is the ligand concentration¹¹². Δt_{RNAP} is defined as the time required for transcription of the riboswitch from the end of the aptamer domain (the earliest possible time at which binding can occur, or $t = 0$) and the uridine-rich tract of the rho-independent terminator stem at the 3' boundary of the expression platform. Assuming the RNA polymerase has a toeprint of 12-14 nucleotides^{153,154}, the aptamer becomes binding competent after the first ~210 nucleotides of transcription (Figure 3.8) and thus the RNAP has to transcribe

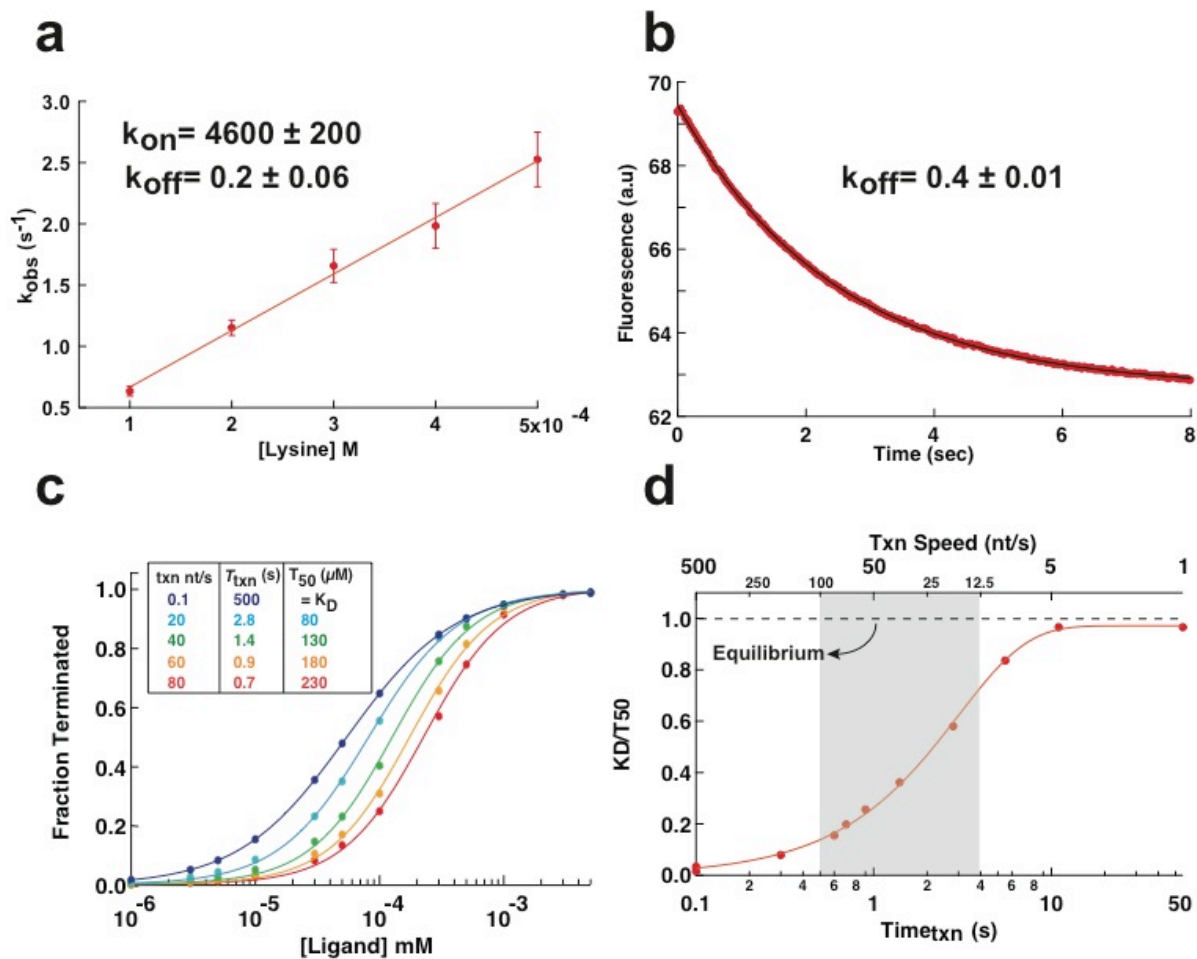


Figure 3.7. Analysis of lysine binding kinetics by stopped flow fluorescence spectroscopy.

(a) Fit to observed binding rate (k_{obs}) as a function of increasing concentrations of lysine. RNA was mixed in a 1:1 volume with increasing ligand concentrations and change in fluorescence as a function of time was measured using a stopped flow fluorimeter. The slope of the line fit yields the association rate constant, while y-intercept of this line yields an estimate of the k_{off} . B) Direct measurements of k_{off} measured by a 1:1 rapid dilution of pre-incubated RNA ligand complexes into buffer. The data shown represents an average of 5 injections. C) The rate constants were used to calculate theoretical estimates of the fraction bound at the times indicated, providing estimates of the T_{50} if the corresponding transcription rates are assumed to be valid. The long time point (dark blue) represents a fully equilibrated binding reaction for comparison. By 5 seconds (light blue), binding is nearly equilibrated. Increasing ligand concentrations are required to elicit termination as transcription speed increases due to a shortening temporal window. d) Visualization of the K_D/T_{50} ratio shows that the reaction is predicted to be fully equilibrated by 10 seconds (as the ratio approaches 1), with appreciable equilibration by 3-5 seconds. The transcription speed axis is provided for additional reference.

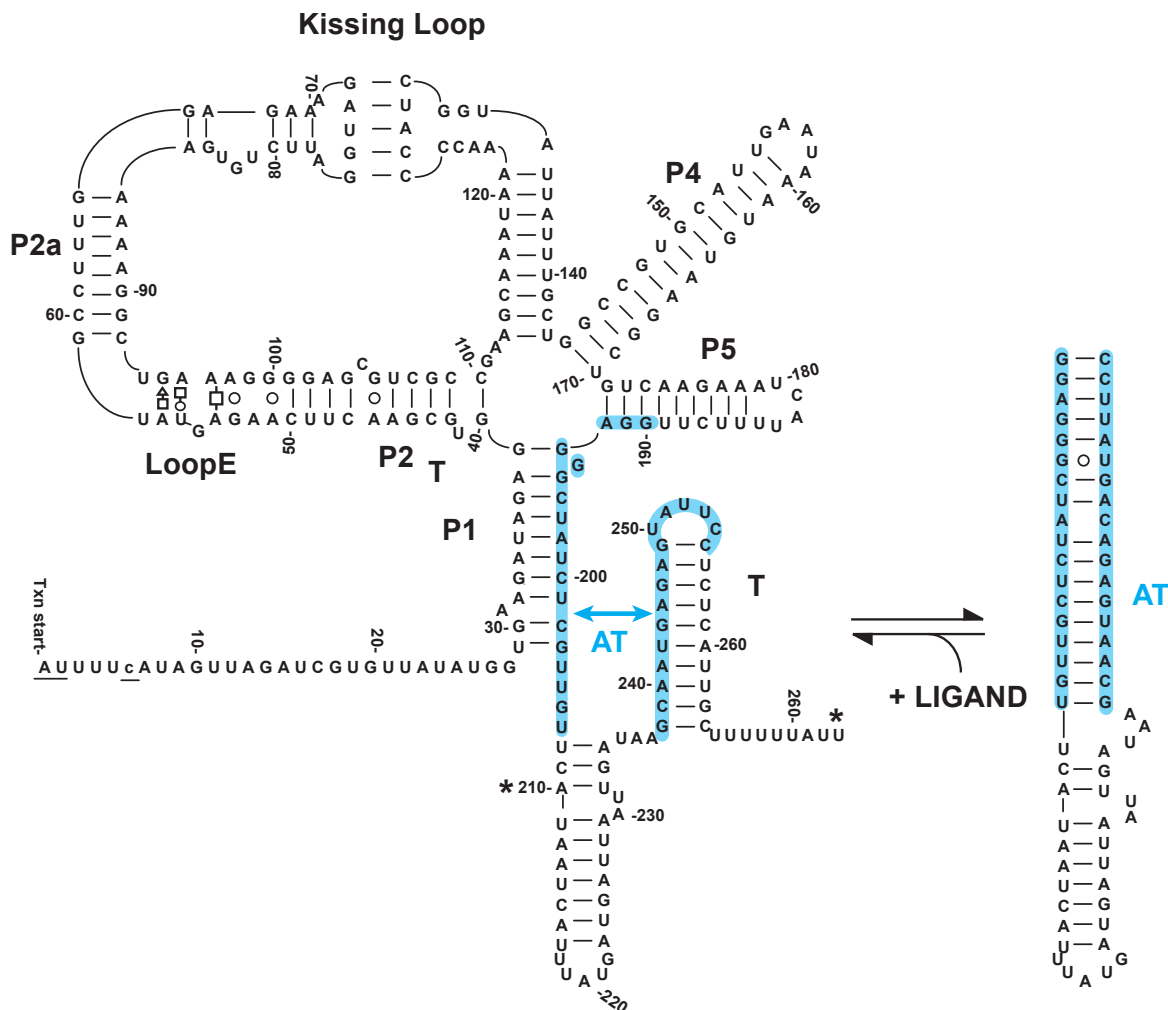


Figure 3.8. Sequences and structure of the *lysC* riboswitch

The 5' leader sequence of the *lysC* riboswitch used in the *in vitro* transcription studies. The transcription templates contained an additional 95 nt downstream of the U-rich tract to differentiate read through from terminated products. Based on the 12-14 nt. footprint of bacterial RNAP) the star at position 210 indicates the point in transcription at which the aptamer would become binding competent. The red star at position 263 represents the putative termination site. This ~50 nucleotide expression platform is used to derive based on the simplifying assumption that transcription proceeds without significant pausing.

an additional ~50-60 nucleotides to reach the downstream termination site. Using physiologically realistic rates of transcription in bacteria (10-80 nt/s)^{155,156}, theoretical T_{50} values were modeled using the above experimental values for the kinetics of lysine binding to the riboswitch, where each individual curve represents a specific rate of transcription (Figure 3.7C). This analysis demonstrates a shift in the regime of ligand sensitivity toward higher concentrations at shorter time intervals. For reference abnormally long time frames demonstrate approach to thermodynamic equilibrium (dark blue). The times correspond to the range of physiologically relevant concentrations. The ratio of the calculated T_{50} and the measured equilibrium dissociation constant for ligand binding to the aptamer ($T_{50,calc}/K_{D,obs}$) plotted against the rate of transcription (or alternatively, Δt_{RNAP}) describes the conditions under which the riboswitch is kinetically versus thermodynamically controlled, where later is defined as $T_{50,calc}/K_{D,obs} = 1$ (Figure 3.7D).

This analysis predicts that the *B. subtilis* *lysC* lysine riboswitch is under thermodynamic control where Δt_{RNAP} of ≥ 3 -5 seconds, corresponding to transcription speeds of ~12-20 nt/s. These rates are consistent with those of the *E. coli* RNA polymerase in the presence of the NusA, a highly abundant cellular transcription factor that slows the translocation step of elongation^{157,158}. On the other hand, if we assume maximal transcription rates of ~80 nt/s, these calculations predict that the T_{50} will increase by a factor of ~5 fold (from ~ 50 μ M to ~250 μ M). Kinetic measurements of a representative set of the lysine analogs (Table 3.1) reveal that most modifications to the ligand slow the association kinetics (k_{on}) by at least an order of magnitude. The most significant drop in the association rate is observed with *N*- ϵ -acetyllysine that abrogates

the electrostatic interaction between the side chain and RNA. Conversely, lysinamide exhibits an ~40-fold increase in the association rate over lysine. One explanation for this observed trend in association rate is that it correlates to the net charge of the ligand; each increase in the net charge by +1 results in substantial increase in k_{on} . Direct dissociation rate (k_{off}) measurements were also made for each compound; these values were more consistent with that of lysine, except for lysinamide which dissociates ~200-fold more rapidly than lysine. The K_D calculated using the observed k_{on} and k_{off} values is within ~ 1-2 fold of that obtained from equilibrium studies for lysine (Table 3.1). Deviations were observed for the ACL and EEL compounds may reflecting a more complex binding mechanism that is not readily apparent from the current analysis.

Table 3.1 Ligand binding kinetics

Ligand	k_{off}	k_{on}	K_{D-calc} (μM)	K_{D-obs} (μM)	K_{D-calc}/K_{D-obs}
Lysine	0.25	4600	54.3	53 ± 6.0	1.0
EEL	0.31	480	625	250 ± 30	2.5
KG	0.55	600	910	570 ± 70	1.6
NML	1.55	690	2200	1670 ± 300	1.3
LysN	48.0	186000	270	90 ± 30	3.0
ACL	0.26	49	5000	1800	2.8

3.5- Assessing the effects of transcription speed on regulation.

To test the predictions made by the above kinetic model, a single turnover *in vitro* transcription assay was used that directly measures the amounts of read through and terminated transcript as a function of ligand. Transcriptions were performed with *E. coli* RNA polymerase at 50 μ M NTP concentrations using a DNA template containing the genomic *B. subtilis* *lysC* leader sequence fused to a T7A1 promoter as described previously (Figure 3.8). While this system does not fully recapitulate the transcription by the native *B. subtilis* RNA polymerase, previous studies of riboswitches suggest that the observed T_{50} is not significantly perturbed by the choice of enzyme^{45,89,109,142}. Indeed a recent study employed the *B. subtilis* polymerase. Notably, protocols outlined for these *in vitro* transcription assays typically exclude potassium from the transcription buffer and utilize high concentrations of Mg^{2+} (20 mM)^{159,160}. The lysine riboswitch is highly sensitive to both the identity and concentration of monovalent and divalent cations⁹⁵, which significantly influences the affinity of the *lysC* aptamer for lysine. To account for this, we have matched the buffer conditions used for equilibrium and kinetic measurements for the 2AP RNA to those of the transcription assay (50 mM Tris-HCl pH 8.0, 5 mM $MgCl_2$, 10 mM KCl and 50 mM NaCl) to ensure an accurate comparison between the results of all experiments.

Transcription at low NTP concentrations (50 μ M) revealed that the apparent T_{50} ($T_{50,50 \mu M}$) response for lysine is approximately equal to the $K_{D,app}$ when the quantified data is fit to a two-state transition model (Figure 3.8, Table 3.3). This observation is consistent with kinetic simulations that predict the riboswitch to be under kinetic control under these conditions. These results suggest that previous differences between the

$K_{D,app}$ and $T_{50,app}$ ^{89,107} can be largely accounted for by the fact that transcription and binding measurements were performed at different temperatures and ionic conditions. The close agreement between the $K_{D,app}$ and $T_{50,app}$ also suggests that folding of the aptamer into a binding competent state occurs on a timescale that does not significantly influence the binding reaction. In support of this hypothesis, polymer models of RNA folding would predict a folding time constant of 0.175 seconds based on the 170 nt chain length of the lysine aptamer domain¹⁶¹.

One issue with the above experiment is that the concentration of NTPs used (50 μ M) is much less than that found within rapidly growing *E. coli*¹⁴⁷. Instead, this condition corresponds to the lower limit of the NTP concentrations found in a quiescent cell^{84,147,162}. To further define the relationship between binding kinetics and regulatory response we measured the T_{50} for lysine over a range of NTP concentrations, which increases the rate of nucleotide incorporation and thereby decreases Δt_{RNAP} (Figure 3.8B). Between 50-400 μ M NTPs we observed no significant change in the T_{50} , however a relatively sharp increase is observed at concentrations of between 400 and 800 μ M, reaching a maximum of ~250 μ M for lysine at high NTP concentrations. This ~5 fold change in the regulatory response agrees surprisingly well with our simplified 2 state model of the binding reaction assuming constant transcription speeds (Figure 3.9B). We also observed that the dynamic range of the regulatory response decreases slightly as a function of transcription speed (Figure 3.9D), suggesting that at rapid transcription speeds the terminator element has insufficient time to form before the RNAP moves past the uridine-rich tract. Similar loss of regulatory efficiency has been observed transcriptional regulation of the *trp* operon by the RNA binding TRAP protein¹⁶³.

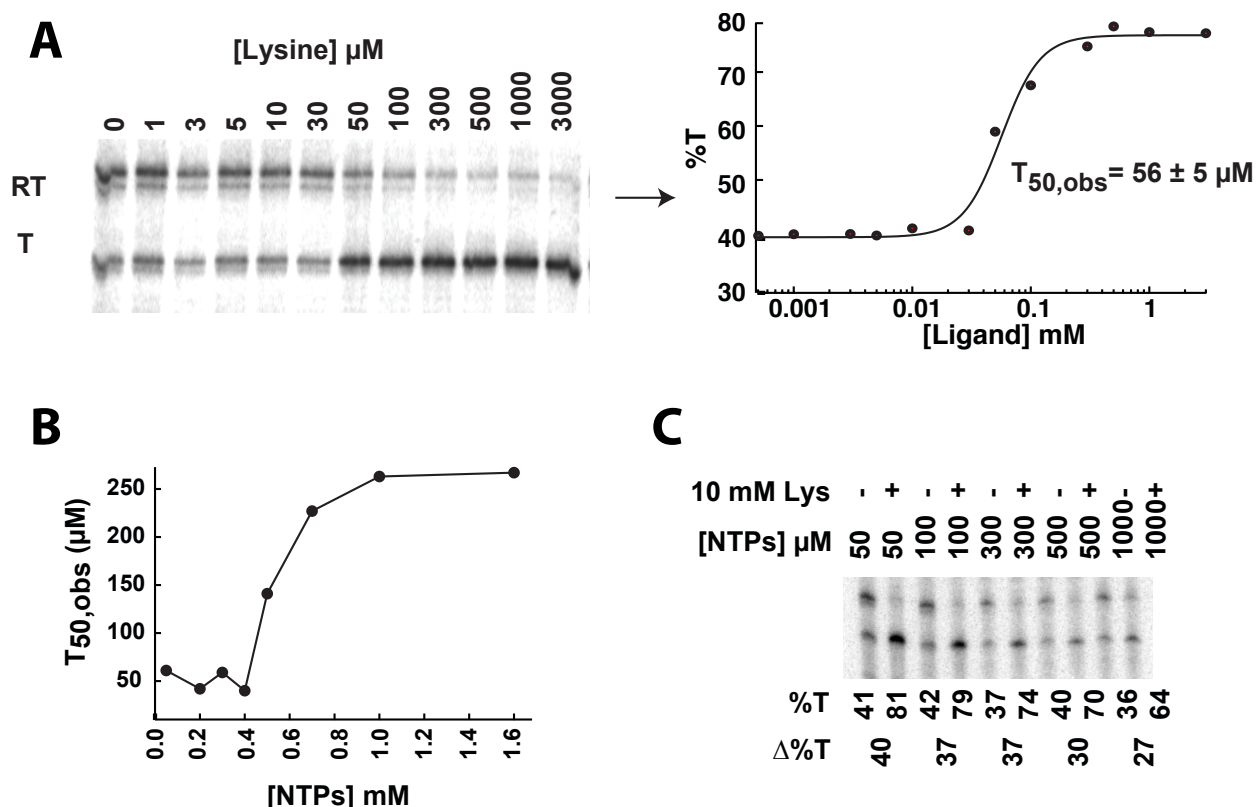


Figure 3.9. Effects of transcription speed and ligand binding kinetics on transcriptional regulation

(a) A representative ligand titration using the *E. coli* polymerase for in-vitro transcription. The bands were quantified as described in the methods and materials to obtain measurements of the $T_{50,app}$. B) Influence of increasing NTP concentrations on the $T_{50,app}$ for lysine. The values indicated on the x-axis represent the concentration of each individual NTP. The non-linear dependence likely denotes that these experiments sample NTP concentrations well below and well above the K_m of the polymerase. The transcription speeds appear to increase rapidly around 400 μM NTPs before leveling off at 1 mM NTP concentrations. Note the 5 fold increase in the $T_{50,app}$ consistent with kinetic predictions. C) The efficiency of the ligand dependent termination response decreases with increasing NTP concentrations. Combined with the increased $T_{50,app}$ effect these data demonstrate ability of transcription speed to modulate ligand dependent regulation.

To assess the effects of the various functional groups on specificity of the regulatory response we tested various lysine analogs at NTP concentrations of 50 μ M ($T_{50\text{slow}}$) and 1 mM ($T_{50\text{fast}}$) (Figure 3.10, Table 3.2). Analogs that perturb recognition of the main chain carboxylate and interfere with the RNA's ability to coordinate the K^+ were found to elicit regulation under both fast and slow transcription rates. Interestingly, while the $T_{50\text{slow}}$ responses were largely consistent with kinetic predictions, the $T_{50\text{fast}}$ values were significantly higher than expected for these compounds. For example, binding kinetics of EEL (and presumably MEL) predict a 4 fold change in the regulation assuming the $T_{50\text{fast}}/T_{50\text{slow}}$ ratio of about ~ 4 , or, similar to that observed for the cognate ligand however loss of the formal positive charge was concurrent with the loss of regulatory activity. Loss of the main chain amine had the most pronounced effect on regulation, as 6-ACA exhibited no apparent activity at any of the transcription conditions tested.

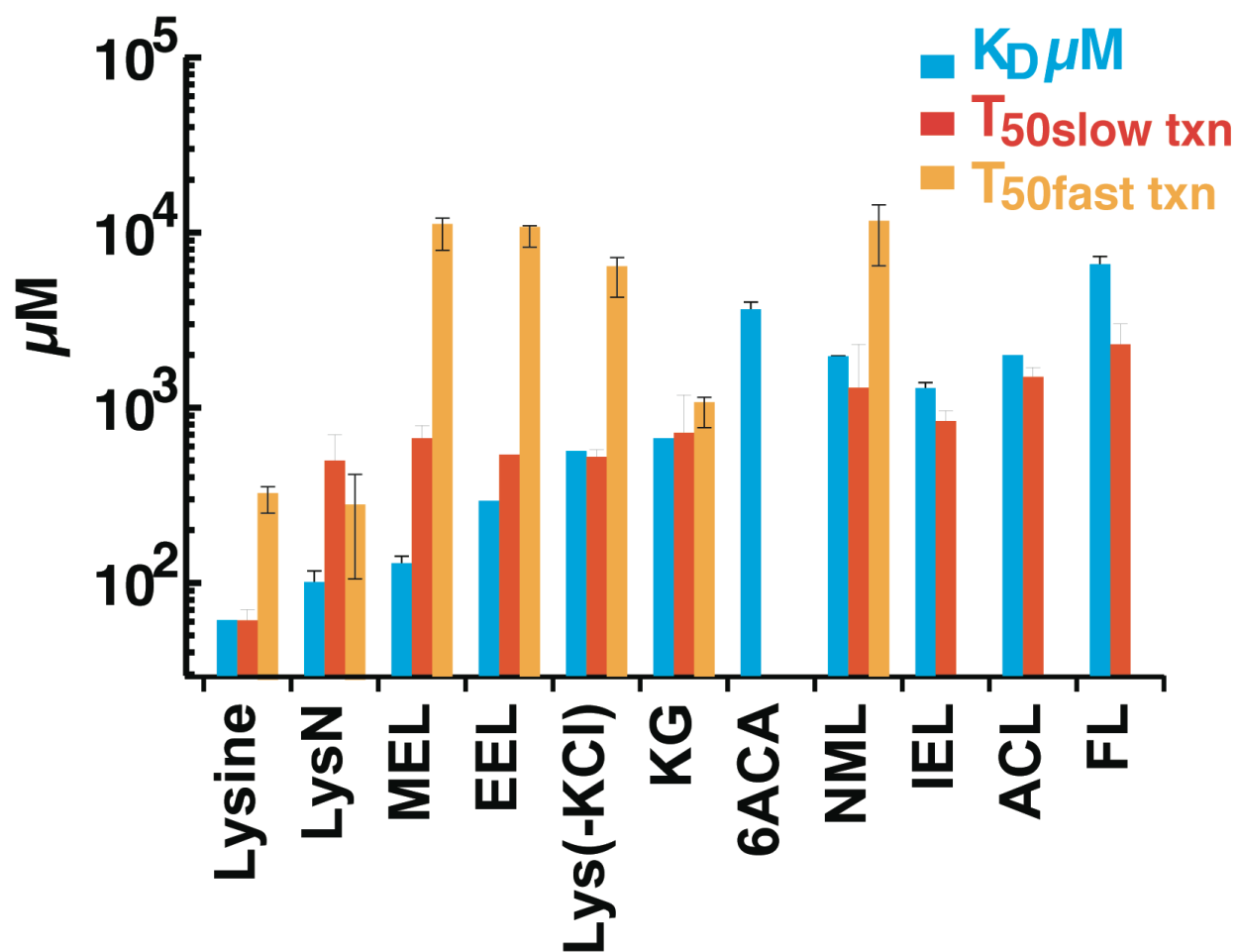


Figure 3.10. K_D and T_{50} comparisons for lysine and lysine analogs.

$K_{D,app}$ measurements (blue) as determined by fluorescence spectroscopy by the 2AP labeled construct. Observed T_{50} at 50 μM NTPs (T_{50slow} red) and 1 mM NTPs (T_{50fast} , orange). Bars are excluded for ligands that were unable to elicit regulation under the conditions indicated. Note the y-axis is depicted on a log scale to allow more effective visualization.

Table 3.2 Comparison of aptamer binding to regulatory control

Ligand	K_D (μ M)	$T_{50-50 \mu M}$ (μ M)	$T_{50-1 mM}$ (μ M)	$K_D / T_{50-50 \mu M}$	$K_D / T_{50-1 mM}$	$\Delta T_{50}(1mM/50\mu M \text{ NTPs})$
Lysine	53 ± 6	61 ± 9	290 ± 50	0.87	0.18	4.8
Lys(-KCl)	480 ± 40	525 ± 50	5500 ± 1400	0.91	0.09	10.5
LysN	90 ± 30	500 ± 200	250 ± 150	0.18	0.36	0.5
MEL	110 ± 30	670 ± 120	9600 ± 2000	0.16	0.01	14.3
EEL	250 ± 30	540 ± 100	9200 ± 1300	0.46	0.03	17.0
KG	570 ± 70	720 ± 460	920 ± 180	0.79	0.62	1.3
6ACA	3100 ± 900	n.d	n.d	n.d	n.d	n.d
NML	1670 ± 300	1300 ± 1000	10000 ± 3800	1.28	0.17	7.7
IEL	1100 ± 290	840 ± 120	n.d.	1.31	n.d.	n.d.
ACL	1700 ± 100	1500 ± 200	n.d.	1.13	n.d.	n.d.
FL	5600 ± 1700	2300 ± 700	n.d.	2.43	n.d.	n.d.

3.5- Discussion

Lysine biosynthesis is an important aspect of normal cellular metabolism beyond its requirement as a component of proteins. For example, key intermediates of lysine biosynthesis such as *meso*-diaminopimalate and dihydropicolinate are essential for cell wall viability and sporulation, placing a strong pressure to accurately control this important metabolic pathway. Previous studies of the lysine riboswitch have hypothesized that specificity of this RNA for L-lysine over related cellular metabolites is primarily mediated through recognition of the positively charged ϵ -amine and its distance from the main chain^{22,89}. In this study, we have further explored this hypothesis

by investigating the ability of the lysine riboswitch aptamer to bind a series of chemical analogs that present specific challenges to recognition of polar functional groups in the ligand.

Our data strongly supports this hypothesis, and further demonstrates that while position of the ϵ -amine is essential, a formal positive charge is not an absolute requirement for binding. This element of recognition is mediated both by a solvent contact that bridges the ϵ -amine and the backbone of G114 as well as an electrostatic interaction between lysine and the RNA. Additional selectivity for lysine appears to be gained by forming relatively weak interactions with main chain atoms thereby prohibiting more promiscuous amino acid binding. Previous studies, which have focused significantly upon the methylene groups of the side chain, as represented by the antimetabolite AEC that binds this riboswitch, showed that the aptamer can accommodate even relatively bulky groups such as a sulfonyl group at the C4 position. Combined with our data on the tolerance of the RNA to bulky modifications of the carboxylate group, it is clear that this RNA can bind a variety of compounds similar to lysine. This is opposed to the RNA's inability to recognize other biological amino acids, which is certainly a primary selection pressure on the aptamer domain. The ability of the lysine riboswitch to productively bind a variety of compounds is somewhat surprising in light of the apparent rigidity of the lysine binding site.

While many studies have examined the thermodynamics and kinetics of ligand binding to riboswitch aptamers this study endeavored to correlate these data with an *in vitro* transcription assay capable of quantifying the regulatory response. To do this, we adopted a kinetic model developed by Breaker and Crothers^{112,164} (eqn. 1) that relates

the regulatory activity (T_{50}) to kinetic binding parameters (k_{on} and k_{off}) and a time allowed for equilibration as determined by the rate of transcription (Δt_{RNAP}). Surprisingly, from this analysis, we observe that at low NTP concentrations lysine dependent regulation is essentially under thermodynamic control, in contrast to previous observations^{89,107}. However, this is consistent with expectations from a calculation of the T_{50} using the experimentally observed binding kinetics and a moderate rate of transcription. The close agreement between K_D and T_{50} for the ligands tested here suggests to us the possibility that some riboswitches that control transcription may have evolved rapid dissociation constants to enable thermodynamic control under cellular conditions that allow slow transcription speeds. A second important observation is that compounds that contain modifications of the carboxylate are most detrimental to the T_{50} ; even at low NTP concentrations, they control the riboswitch in a kinetic fashion. We speculate that this may be because this region of lysine directly contacts the “switching sequence” (Figure 1a) that is directly involved in regulatory structural switch.

It is important to note that our kinetic model does not account for some important details of the transcriptional response. For example, 30-40% of the transcripts are terminated in the absence of ligand and reach an upper limit of ~80% termination under saturating ligand concentrations. This observation is typical of the regulatory response of many riboswitches, and reflects aspects of the cotranscriptional folding process that are not ligand dependent, and therefore excluded from our binding model^{116,165,166}. A second assumption that we make is a direct relationship between the sequence length and Δt_{RNAP} , which implies that transcription proceeds at a constant rate. This fails to account for the potential for transcriptional pausing during synthesis of the *lysC*

expression platform, as observed in the *B. subtilis* FMN and *E. coli* B12 riboswitches^{109,167}. Such a phenomenon may explain why the binding kinetics of non-cognate ligands do not accurately predict the regulatory response under rapid transcription conditions. The difficulty of identifying pause sites from the primary sequence alone¹⁶⁸, motivates further empirical examination of this phenomenon in the context of the lysine riboswitch and other riboswitch families.

The influence of NTP concentration on transcription attenuation has been studied for a number of systems including protein dependent regulation of the *trp* operon^{163,169}, and tRNA dependent regulation of the *glyQS* gene¹⁴². As was observed for the *lysC* riboswitch (Figure 3.9 and 3.10) regulation by the cognate effector is markedly influenced at similar NTP concentrations (400-500 μ M) *in vitro*. As transcription attenuation represents a broadly utilized regulatory mechanism for biosynthetic operons in bacteria¹⁷⁰ it is tempting to speculate that a collective increase in the sensitivity of attenuation mechanisms may provide a global regulatory strategy for optimizing energy consumption under resource limited conditions. Such a strategy would presumably complement other NTP sensing mechanisms such as those that control pyrimidine biosynthesis¹⁷¹ and ribosome biosynthesis¹⁷², the latter of which is a key determinant of the overall growth rate of bacteria. On the other hand the ability of the lysine riboswitch to integrate a more global picture of metabolism into its regulatory output may reflect a specific selective pressure to tune this element based on the central role of lysine biosynthesis in bacterial growth. Further analysis of riboswitch mediated transcription attenuation will be therefore required to decipher the degree to which different riboswitches couple regulation to global aspects of metabolism, and should yield

additional insights into how these organisms tailor their metabolism to function robustly in diverse environmental conditions.

3.6- Methods and Materials

Preparation of labeled RNA

A 2-aminopurine (2AP) modified oligonucleotide with the sequence GGAGUCUUUCUUGGAG-2AP-GCUAUCUCUCC was chemically synthesized by Dharmacon Inc. All other RNAs in this study were synthesized using T7 polymerase and purified according to previously established protocols⁶³. Purified RNAs were stored in 5 mM Tris-HCl pH 8.0, 0.5 mM EDTA buffer and stored at -20°C prior to use.

To assure proper annealing of the two piece RNA construct the sequence derived from the *B. subtilis* lysC aptamer was engineered using the following 5' and 3' primers to PCR amplify from a cloned plasmid containing the entire promoter and 5' leader segments of this gene:

Forward primer-

TAATACGACTACTATAGGCGTGAGCAGACTCTTTTTTGGAGAGATAGAGGTGCGAAC

Reverse primer 3'-

TAAACGACCGGCACGTAACCTATTACATTCCGACAGTTCTTTCTGAGG

This approach generated an RNA with the sequence:

GGAGAGAUAGAGGUGCGAACUUCAAGAGUAUGCCUUUGGAGAAAGAUGGAUUC
UGUGAAAAAGGCUGAAAGGGGAGCGUCGCCGAAGCAAUAAAACCCCAUCGGUUAU
UAUUUGCUGGCCGUGCAUUGAAUAAAUGUAAGGCUGUCAAGAAAGACUCC.

The two-piece RNA used in the titration experiments was constructed by annealing in a 50 mM HEPES pH 8, 100 mM NaCl buffer. The unlabeled RNA was added to a final concentration of 3 μ M to the a 1 μ M stock of 2AP labeled RNA and heated in a PCR block to 85°C for 2 min, and cooling at a rate of 0.1°C/sec with a 2 min hold at 10° intervals down to 4°C. Annealed products were analyzed by native polyacrylamide gel electrophoresis to ensure the quality of the annealing reaction was optimal (data not shown). Under these conditions the labeled oligo was verified to incorporate with high efficiency into a single species corresponding to the appropriate size product.

Fluorescence Spectroscopy

All ligand titrations were carried out in 50 mM Tris-HCl pH 8.0, 100 mM NaCl and 5 mM MgCl₂ and 10 mM KCl with a 100 nM RNA unless indicated otherwise. Changes in the fluorescence of the 2AP reporter as a function of ligand concentration were measured using Tecan 96 well plate reader for all room temperature experiments and in a Chirascan fluorescence spectrometer for 37°C experiments. Data was fit to the equation:

$$\text{Fraction Bound} = \frac{F_{\max} - F_{\min}}{F_{\max} - F_{\min}} \left(\frac{[L]^n}{[L]^n + K_D} \right)$$

Where ΔF is the measured change in fluorescence [L] is the ligand concentration, and $F_{\min}$ is the initial fluorescence and $F_{\max} - F_{\min}$ represents the total magnitude of the change in measured fluorescence change for the titration.

Isothermal Titration Calorimetry

The wild type lysC riboswitch aptamer from *B. subtilis* (RFAM accession ABQL01000005.1/1891692-1891513) was dialyzed overnight at 4°C in 50 mM HEPES pH 8.0, 100 mM NaCl, 5 mM MgCl₂, and 10 mM KCl. The RNA was diluted to a final concentration of 60 μ M and titrated with lysine that had been dissolved directly in the dialysis buffer at concentrations 10 fold in excess of the RNA. Titrations were all performed at 20° C and 37°C using the VP-ITC micro calorimeter (Microcal, Inc.). Data analysis and fitting was performed with the Origin 5.0 software suite (Origin Laboratories) as previously described ¹.

Chemical probing

Chemical probing was performed as described elsewhere^{79,80} with the following adjustments. RNA was diluted to 100 nM in 10 μ L of a buffer of 50 mM Tris-HCl pH 8.0, 100 mM NaCl, 50 mM KCl, 5 mM MgCl₂ and 0.1 mM EDTA. The RNA was incubated with either 1 mM lysine or 10 mM of a lysine analog to ensure binding was at or near saturation. RNA was probed at 37°C for 45 min (5~ half lives) with 65 mM NMIA. Reverse transcription was performed by adding 3 μ L of ³²P labeled DNA primer (~10 nM final) and heating the solution to 95°C for 5 min then cooling to 37°C for 5 min on a PCR block. The solution was then brought to 54°C before adding 3 μ L of a Superscript III RT mix and allowing 10 min for primer extension. The RNA was then degraded by adding 1 μ L of 4 M NaOH and incubating at 95°C for 5 min before quenching with a 1:1 Tris-HCl(unbuffered):formamide solution. PAGE analysis of the remaining cDNA was performed using 12% acrylamide gels.

In vitro transcription studies

In vitro transcription assays were performed using the wild type lysC gene from *B. subtilis* containing the T7A1 phage promoter fused at the genomic +1 transcription start site. Template DNA (10 nM) was equilibrated at 37°C in the presence of 0.5 units of *E. coli* RNA polymerase holoenzyme with saturating α -subunit (Epicentre) in 1X TB1 (70 mM Tris-HCl pH 8.0, 20 mM NaCl, 5 mM MgCl₂, 10 mM KCl, 17 mM DTT,.1 mM EDTA, 35 μ g/mL BSA,) for 10 min prior to the addition of NTP substrate to 50 μ M in a final volume of 20 μ L. Reactions were incubated an additional 15 min at 37°C and quenched with an equal volume of 85% formamide. Products were separated by 6% denaturing gel electrophoresis, and counts obtained by phosphorimaging. Total counts for full length and terminated were background corrected and normalized to the total number of A's in each product. Percent termination was plotted and fit to the equation

Crystallographic data collection and processing

Crystallization of the 6ACA and KG ligands was performed with a variant of the *Thermatoga maritima* lysine riboswitch that was previously described (PDB ID: 3D0U)¹⁴³ RNA at 400 μ M was pre-equilibrated with saturating amounts of each ligand and mixed in a 1:1 ratio with mother liquor consisting of 2 M Li_2SO_4 and 5 mM MgCl_2 . Diffraction data was collected on a copper rotating anode source (Rigaku RU200 and Rigaku RU2HR) with a Rigaku MSC IV++ area detector at 100 K. Data were scaled and averaged using the D*TREK as part of the CrystalClear software package¹⁷³(Rigaku). Structures of the 6ACA and KG bound complexes were solved by molecular replacement with the unliganded lysine aptamer (PDB ID: 3D0X) using the PHENIX software suite¹⁷⁴Model building was performed with Coot¹²⁶ and additional refinement was performed with the routines implemented in PHENIX.

Chapter 4: Rational design and optimization of chimeric riboswitches

One of the central aims of synthetic biology is to extract interchangeable parts from nature that can be reassembled to create novel chemical systems^{175,176}. RNA provides a particularly attractive tool for this purpose as it plays diverse regulatory roles in nature, and forms structures with predictable stabilities thereby enabling forward engineering of many desired functional parameters^{177,178}. Numerous studies have demonstrated the utility of RNA design by constructing RNA devices that mimic the regulatory features of natural sRNAs¹⁰. These devices promote changes in gene expression through the formation of RNA-RNA interactions, and afford modulation of the gene expression response based on designed stabilities of the RNA hairpin and duplex structures¹⁷⁹⁻¹⁸². Such elements have been employed to program novel cellular behaviors such as inducible cell death¹⁸⁰ or cellular circuits that can perform simple counting functions¹⁸³.

While RNA-RNA based devices offer many advantages from a control standpoint, they are unable to directly sense the chemical changes in the cell that are often of central interest and importance. The ability of nature to produce RNAs that adopt complicated structures to perform catalysis or binding however has been replicated by *in vitro* selection methods or “SELEX”¹⁸⁴ to generate aptamers for a wide variety of proteins¹⁵ and natural or synthetic small molecules¹⁸⁵. The key challenge for employing SELEX aptamers in the design of ligand sensing devices has been to identify efficient ways of coupling the aptamers to regulatory domains that can readily interface with the cellular machinery¹⁸⁶. The most common approach has been to create so called

“aptazymes”, or aptamers fused to ribozymes, that trigger mRNA cleavage in a ligand dependent fashion by coupling ligand binding to stabilization of the ribozyme cleavage site¹⁸⁷⁻¹⁸⁹. This strategy can not only be used to generate standard ON/OFF switches, but by coupling multiple ligand binding domains to a single ribozyme scaffold one can construct more sophisticated “AND,” “NOR,” and “NAND”¹⁹⁰. On the other hand aptazymes often require regulatory domains that can significantly destabilize the aptamer domain to achieve a detectable expression response, thereby limiting the ability to tune the ligand sensitivity of these devices.

As an alternative to the aptazyme strategy another study recently demonstrated that a translational regulatory domain could be selected for *in vivo*¹⁹¹. Using a SELEX derived aptamer for atrazine, a commonly used herbicide that is toxic to humans, this creative approach selected for RNAs that could drive expression of a motility gene in the presence of atrazine. Cells carrying sequences that encode functional regulatory RNAs could thereby be spatially isolated from their non-functional counterparts on agar media. As with aptazymes however, the addition of the regulatory domain appeared to compromise the the function of the aptamer: in this case ligand concentrations in ~1000 fold excess of the aptamer K_D are required to achieve efficient regulation, though this could also reflect changes in the bioavailability in cells¹⁹¹.

Riboswitches like those commonly found in the 5' leader sequence of bacterial mRNA have provided insights into the diverse design strategies that have been adopted by nature, which include regulation at the levels of transcription attenuation, translation initiation, and splicing (Figure 1.3)³⁶. The majority of naturally occurring riboswitches however control gene expression at the level of transcription, by directing the formation

of transcription termination signals in the nascent mRNA. Interestingly, the expression platforms from different riboswitch classes found in the *B. subtilis* genome appear to contain the full complement of structural features necessary for directing regulation outside the context of their native aptamer (Figure 4.1). For example, the sequences that make up the 3' side of the P1 helix can form complementary interactions with the 5' shoulder of the terminator stem to prevent transcription attenuation in the absence of ligand (Figure 4.1 gray sequences). It therefore stands to reason that the regulatory domains of natural riboswitches may exhibit a significant degree of modularity as regulatory switch domains making them ideal synthetic parts that could be reprogrammed to respond to different metabolites by replacement of the native aptamer. Unlike the design strategies described above, which often require selection for the regulatory function, this approach would presumably enable a simple mix and match strategy for engineering novel riboregulatory elements.

Pablo has been largely responsible for demonstrating that this mix and match approach offers a highly reproducible and robust methodology for installing both natural and SELEX derived aptamers into the context of different regulatory domains to obtain functional ligand sensing devices. Joan Marcano has demonstrated this design strategy in-vitro. My role in this effort has primarily focused on understanding how the observed regulatory responses are shaped by structural features of the expression platform, with the goal of outlining the design principles to be considered for obtaining desired response characteristics for future engineering endeavors.

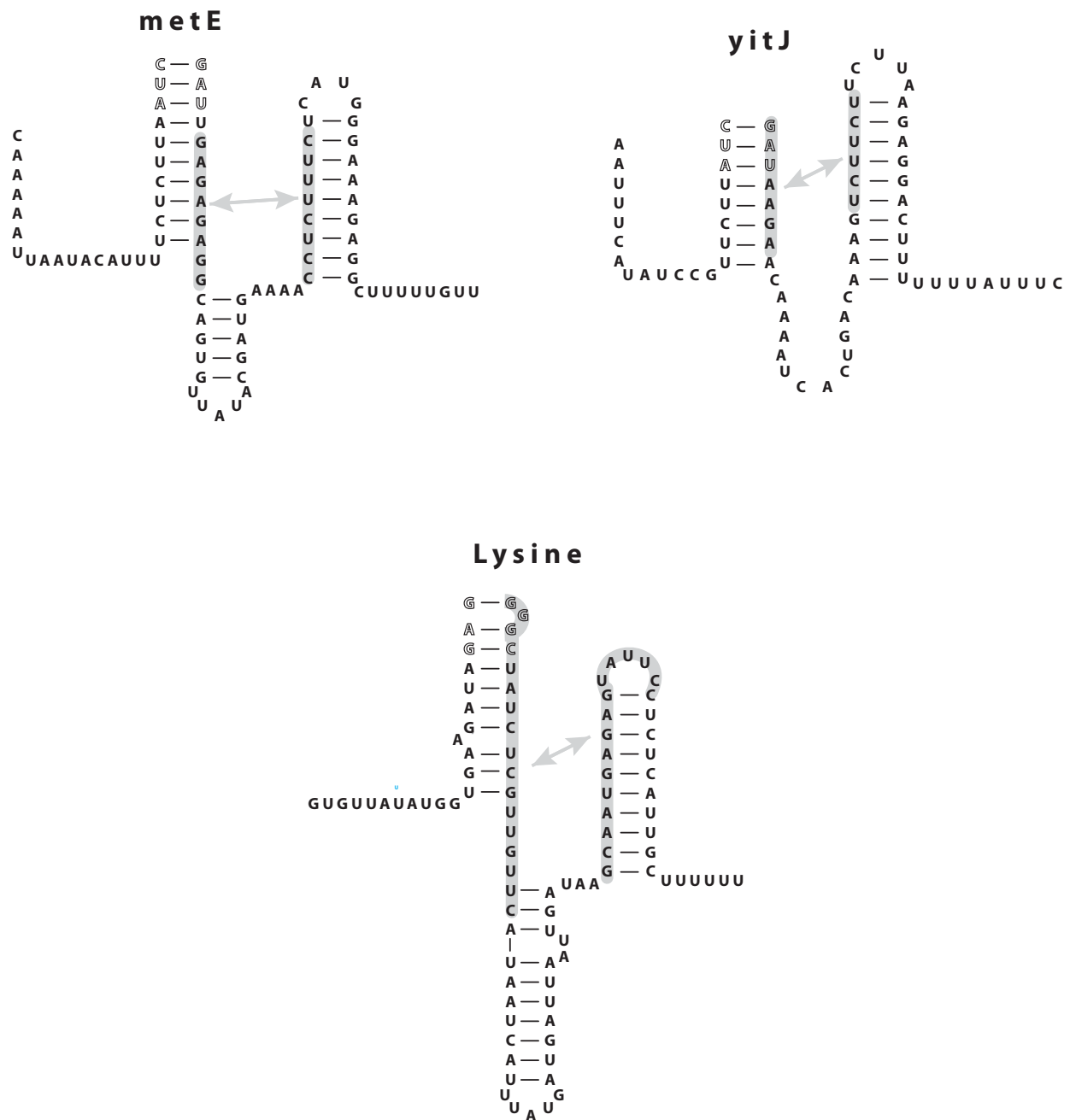


Figure 4.1 Apparent modularity of naturally occurring expression platforms.

Proposed structures of the *metE*, *yitJ*, and *lysC* expression platforms. The sequences involved in the conformational switch are highlighted by in gray. The nucleotides within the aptamer that are highly conserved and required for ligand binding and specificity are outlined for clarity.

4.1- Design strategy for generating chimeric riboswitches

To assess the viability of using natural expression platforms as regulatory scaffolds Pablo Ceres and I have created chimeric regulatory devices in which the native aptamer is replaced by a series of naturally or non-naturally derived aptamers (Figure 4.2). The basic strategy requires identification of the residues that are requisite for ligand binding by the aptamer, and those required for driving a conformational switch by the expression platform (Figure 4.2). Crystallographic structures of riboswitch aptamers and secondary structural models of each expression platform (Figure 4.2) provide the necessary information for choosing the appropriate structural boundaries for constructing these RNAs. Our examination focused on the modularity of expression platforms derived the *metE*, *yitJ* or *lysC* leader sequences in the *B. subtilis* genome (Figure 4.1) as a number of previous studies had demonstrated robust *in vitro* transcription responses for these RNAs and the requisite structural switches could be readily identified. To demonstrate the generality of this approach we selected a variety of natural (SAM, guanine, adenine, lysine, FMN) and SELEX derived aptamers (theophylline and tetracycline) that were constructed using standard recombinant PCR methods to generate full length templates for use in single turnover *in vitro* transcription assays.

4.2- Wild type riboswitches as a benchmark for comparative analysis

To provide a standard for comparative analysis of the fusion constructs we first wanted to develop a consistent model of the native riboswitch systems. As discussed in the previous chapter, transcription can heavily bias the regulatory response by

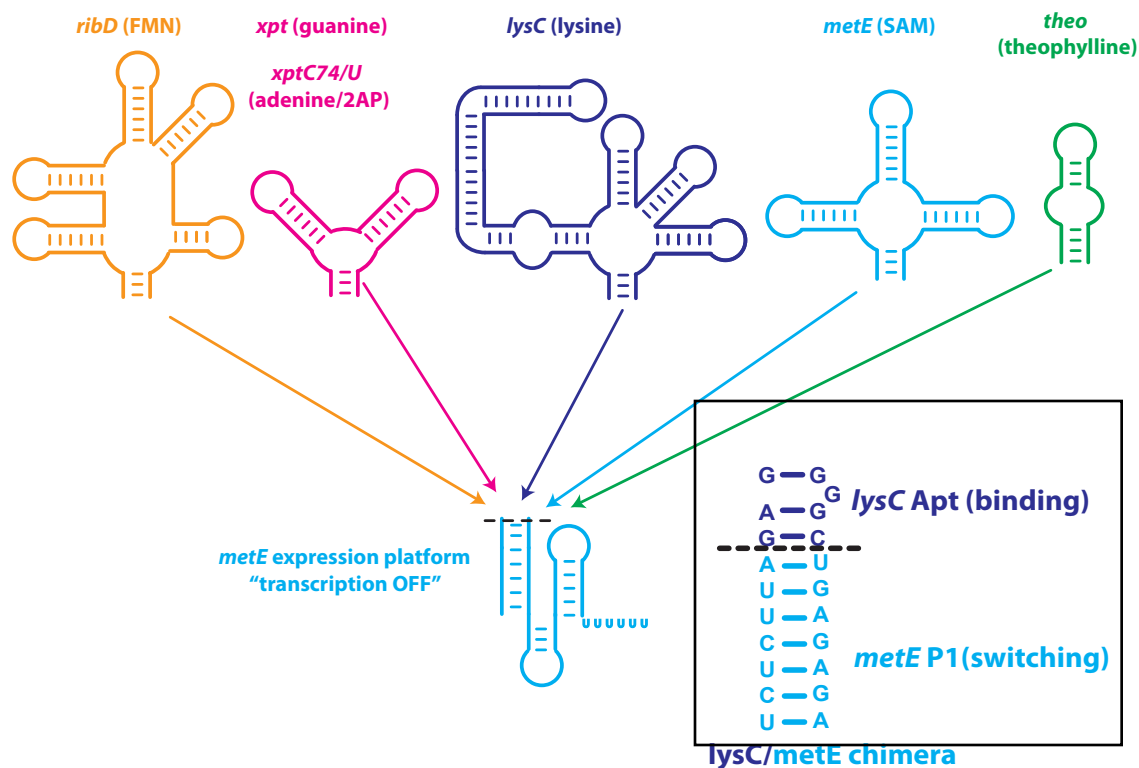


Figure 4.2 Outline of chimeric riboswitch design strategy.

Fusion constructs were generated by making chimeric RNAs that append the riboswitch aptamer into the context of a natural expression platform (*metE*). The various natural and SELEX derived aptamers used in constructing chimeras are diagramed at the top (the tetracycline aptamer was also utilized, but not shown). The abbreviated names indicate the gene from which the sequence was derived. The boxed inset illustrates the rules that were followed in constructing these fusions by fusing aptamers to the P1 helix in a fashion that includes nucleotides necessary for ligand binding from the aptamer and nucleotides important for switching from the expression platform. The aptamer/expression platform formalism (*lysC/metE*) is shown for clarity in the following discussion.

preventing ligand binding from reaching equilibrium, leading to elevated ligand requirements for regulation. My studies of the *lysC* riboswitch however suggested that the differences between the T_{50} and K_D for this RNA could be largely accounted for by the failure of previous studies to match the Mg^{2+} and K^+ ion concentrations between binding and transcription studies, as these ions significantly influence the binding affinity of the aptamer (see Chapter 3). Published K_D values of other riboswitch aptamers¹⁹² have also been largely determined under conditions that significantly deviate from those used in transcription, which could also lead to artificial deviations. Motivated by these discrepancies my analysis of the aptamer binding affinities was therefore conducted under the conditions used for the transcriptional analyses to ensure accurate comparisons could be made (see methods).

Surprisingly, the K_D values obtained by ITC agree remarkably well with the T_{50} values that Pablo obtained for the native riboswitches which were determined using single turnover transcription by *E. coli* RNA polymerase at 50 μ M NTP concentrations (Table 4.1); contrasting with the view that kinetic control may dominate for most riboswitches^{117,193}. Such a conclusion would suggest that the selective pressures associated with transcriptional control may have placed a premium on natural aptamers to equilibrate rapidly in order to effectively sense their environment in a short temporal window. Interestingly, the selection pressures provided by most column based methods would presumably place an emphasis on slow dissociation rates that in turn necessitate long equilibration time, suggesting that these aptamers should be expected to exhibit significant kinetic control. On the other hand the close agreement could reflect the low

NTP concentrations that were utilized for transcription (50 μM), as this can significantly influence the regulatory response.

Table 4.1 Functional parameters of wild type riboswitches *in vitro*

Ligand	Aptamer	K_D , (μM)	ΔG° binding (kcal/mol)	T_{50} (μM)
SAM	metE	2.0 ± 0.5	8.1	2 ± 0.5
SAM	yitJ	1.5 ± 0.3	8.2	2 ± 0.5
lysine	lysC	18 ± 2	6.0	60 ± 5
guanine	xpt	0.024 ± 0.003	10.8	n/a
2-AP*	xpt(C74U)	8.8 ± 0.8	7.1	n/a
FMN	ribD	0.91 ± 0.1	8.5	0.2
theophylline	theophylline*	20 ± 6	6.6	n/a

4.3- Functional characteristics of chimeric riboswitches

Fusion constructs were tested for their ability to drive ligand dependent transcription termination using the same procedures described for the wild type RNAs. The chimeric design strategy proved highly robust as all of the fusion constructs exhibited reliable regulation in the context of the SAM and lysine derived expression platforms, despite distinct differences in the length and structural features of the of these domains (Figure 4.3). The observed T_{50} for the fusion RNAs are for the most part

consistent with the aptamer K_D 's (Figure 4.3A) with some notable exceptions. For example the *yitJ* expression platform appears to provide the best context for placing the *xptC74/U* and *theophylline* aptamers to regulate at ligand concentrations that closely approximate the observed K_D . Conversely, *ribD*, *lysC* and *metE* aptamers require ~10 fold higher ligand concentrations to operate in the context of the *yitJ* expression platform. Similarly of the constructs placed into the *lysC* and *metE* expression platforms only the *ribD/metE*, *theo/metE* and *theo/lysC* constructs were found to significantly deviate from the aptamer K_D under these conditions (Figure 4.3, asterisks). These data imply that the ligand dependent response can be differentially tuned by the context of the expression platform. As the aptamer and expression domains have coevolved in natural systems many parameters could thereby contribute to forming the regulatory response that are not readily apparent from this analysis.

Table 4.2 Summary of chimeric riboswitch data

Ligand	aptamer	K_D , (μ M)	Expression platform					
			<i>B. sub metE</i>		<i>B. sub yitJ</i>		<i>B. sub lysC</i>	
			T_{50} , (μ M)	DR	T_{50} , (μ M)	DR _{GE}	T_{50} , (μ M)	DR _{GE}
SAM	metE	2.0 ± 0.5^a	2	12-95	15	40-85	2.5	48-95
SAM	yitJ	1.5 ± 0.3^a	2	59-97	2	12-95	0.5	35-97
lysine	lysC	18 ± 2^a	160 ^a	31-54	180 ^a	45-69	61 ^a	26-78
guanine	xpt	0.024 ± 0.003^a	0.03	45-92	0.071	32-87	0.073	11-83
2-AP*	xpt(C74U)	8.8 ± 0.8^a	30	62-96	7.3	15-72	32	35-87
FMN	ribD	0.91 ± 0.1^a	8.8	26-72	19	9-51	1.3	10-53
theophylline	SELEX	20 ± 6^a	130	32-91	12.5	35-93	275	5-82

^a Indicates my contributions to this work

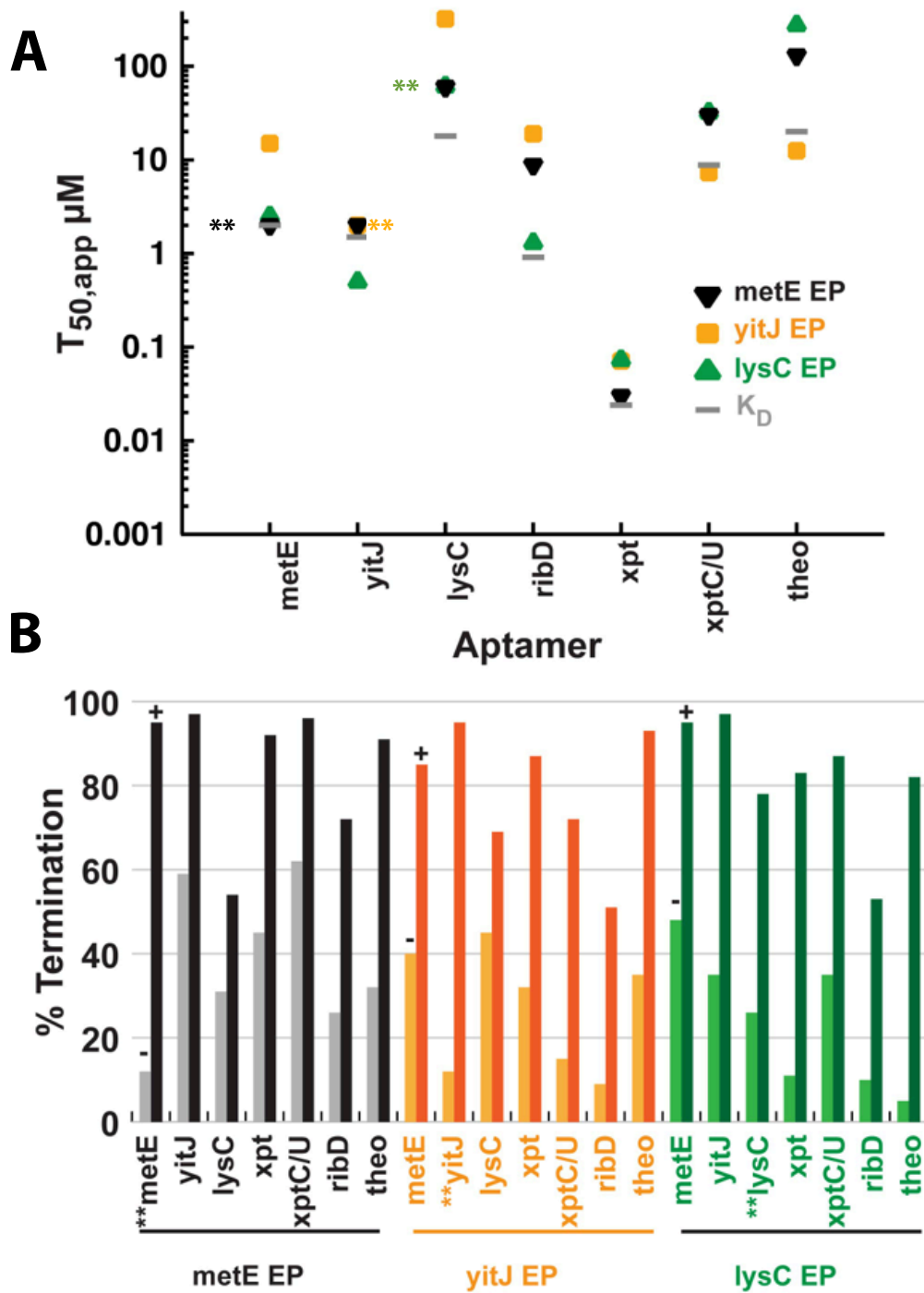


Figure 4.3 Response characteristics of reprogrammed riboswitches.

A) Plot of the observed T_{50} values for the various fusion constructs. X-axis denotes the identity of the aptamer domain B) Plot of the %Termination based of fusion constructs. **Indicates native RNAs for reference

The other important regulatory parameter that is pertinent to these studies is dynamic range of the expression response (DR_{GE}), which corresponds to the change in the fraction of terminated transcripts. The DR_{GE} can therefore be related to the data by the following equation:

$$\frac{\% T (-ligand)}{\%T (+ligand)} = \text{Fold Change}$$

Analysis of the fraction terminated data provides some interesting trends (Figure 4.3B, Table 4.2). For example the 7.9 DR_{GE} of wild type *metE* riboswitch is diminished to 2.1 fold in the *metE/yitJ* chimera and 2.0 in the *metE/lysC* chimera. Likewise the *lysC/metE* and *lysC/yitJ* constructs exhibit diminished response ranges of 1.6 and 1.5 respectively, compared with the *lysC* wild type RNA. On the other hand many fusions exhibit DR_{GE} of greater than 5-fold suggesting a relatively robust readout, with the *yitJ/lysC* fusion behaving much like the wild type *yitJ* sequence. It should be noted however that the DR_{GE} however may not factor as critically in obtaining a strong *in vivo* response as in the absolute termination efficiency. For example the *ribD/yitJ* construct may serve as a poor *in vivo* reporter due to leaky expression in the presence of saturating ligand concentrations, where as many of the fusions into the *metE* expression platform approach almost complete termination that would be better suited for controlling reporter gene expression.

Table 4.3 Expression response range of riboswitch chimeras

Construct (Apt/EP)	Fold Change
metE/metE	7.9
yitJ/metE	1.6
lysC/metE	1.6
xpt/metE	2.0
xpt(C74U)/metE	1.5
ribD/metE	2.8
theo/metE	2.8
metE/yitJ	2.1
yitJ/yitJ	7.9
lysC/yitJ	1.5
xpt/yitJ	2.7
xpt(C74U)/yitJ	4.8
ribD/yitJ	5.7
theo/yitJ	2.7
metE/lysC	2.0
yitJ/lysC	2.8
lysC/lysC	3.0
xpt/lysC	7.5
xpt(C74U)/lysC	2.5
ribD/lysC	5.3
theo/lysC	16.4

4.4- Mutational analysis of the P1 element

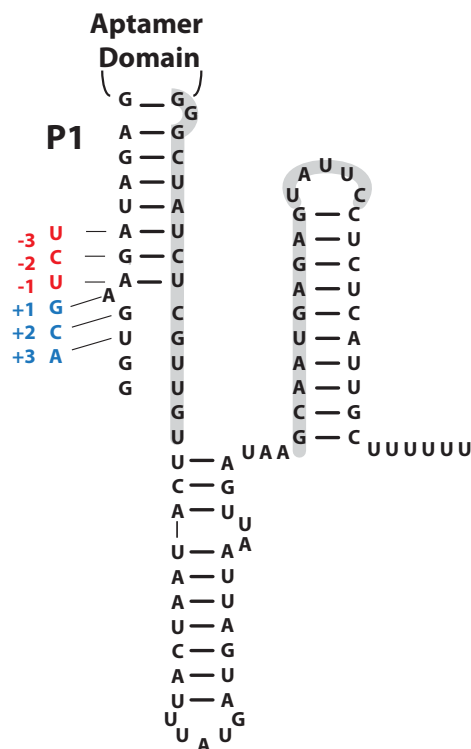
Structural analyses of many riboswitch classes have suggested a common mechanism of regulatory control that involves stabilization of the P1 helix¹⁹⁴ based on the observations that ligand binding typically nucleates contacts that extend the base pairing of the P1 element or introduce tertiary structures that buttress it against strand invasion by the expression platform. In support of this hypothesis optical trapping studies have provided direct measurements of ligand dependent stabilization of P1 for adenine⁸⁷ and TPP¹⁹⁵ binding aptamers. Indeed our basic design strategy is based on

this principle as we have chosen to couple aptamers through the P1 element. These elements however vary amongst the expression platforms utilized in this study that could explain some of the context dependent changes of the regulatory response as mentioned above.

To better understand how alterations to the P1 helix influence the regulatory response I made a series of mutations to the 5' side of the P1 helix of the native *lysC* riboswitch that progressively stabilize or destabilize this element and tested their effects on transcription (Figure 4.4). Extension of the Watson-Crick complementarity by an additional 3 base pairs is predicted to stabilize the P1 by ~ 7 kcal/mol (P1+3, Figure 4.4B) based on predictions using the RNA cofold program as part^{196,197}. This mutation was found to disrupt formation of the mutually exclusive antiterminator structure leading to high levels of termination independent of lysine (Figure 4.4B, P1+3). Thus while the antiterminator (AT) is predicted form a hairpin structure with a ΔG° of -17.7 kcal/mol, the strand invasion process that leads to formation of this element occurs to prevent nucleation and folding of the downstream terminator stem. This is consistent with studies on the the influence of transcription polarity and local structural stability on the folding process which have demonstrated that co-transcriptional folding can produce a folding pathway that is inaccessible by traditional refolding methods^{114,198,199}. A more modest stabilization of ~ 3 kcal/mol (P1+2, Figure 4.4B) to the P1 stem does not strongly influence the DR_{GE} , nor does it impart a significant change to the T_{50} (data not shown), though termination efficiencies are slightly increased.

Destabilization of the P1 element produces the opposing effect on transcription. A loss of ~ 4 kcal/mol (P1-2, Figure 4.4B) severely impairs the ability of the RNA to

A



B

ΔG (kcal/mol) ^o	-3.6		-4.3		-8.1		-8.4		-10.3		-11.8		-15.1	
P1 variant	-3		-2		-1		wt		+1		+2		+3	
Lysine	-	+	-	+	-	+	-	+	-	+	-	+	-	+
%T	47	45	17	28	29	71	33	78	37	81	47	83	93	95
Δ%T	2		11		42		45		44		35		2	

Figure

4.4 Analysis of P1 stability on the regulatory response of the *lysC* riboswitch

A) Schematic demonstrating the positions mutated in this analysis to disrupt or extend perfect base pairing complementarity in the RNA. B) results of transcription in the absence or presence of 10 mM lysine. See text for details of the ΔG^o calculations.

promote transcription termination, leading to the production of primarily read through products with only a slight ligand dependent signal. As crystal structures of the lysine riboswitch would not predict a significant influence for P1 mutations on the binding affinity of the aptamer, this data argues that the P1 element serves as the crux of allosteric communication between the domains, and help to explain the observed conservation of the stem lengths for many natural riboswitch classes³⁷. Furthermore these data imply that changes in $\pm 3-4$ kcal/mol to the P1 stability appear to support a relatively robust regulatory response, explaining the flexibility of the chimeric design strategy. Interestingly decoupling of the binding and folding reactions appears to occur in a relatively narrow energetic window that is near the free energy measured for lysine binding (-6.0 kcal/mol). These data also imply that more refined mutational strategies can be employed to manipulate the DR_{GE} to produce user defined functionality.

To demonstrate the generality of manipulation of the P1 helix Pablo performed a similar set of mutations in the *metE* riboswitch (Figure 4.5A). The P1 stability of the *metE* RNA exhibited the same basic trends as the *lysC* riboswitch however this RNA appears to promote termination more efficiently. For example, despite large energetic penalties associated with the P1-1 and P1-2 mutations the *metE* riboswitch promotes a reasonable degree of termination. Stabilization of the P1 stem by only 2.5 kcal/mol leads to almost complete transition to terminated products.

The importance of variation of the length of the P1 helix in the design of chimeras from synthetically derived aptamers was illustrated for the tetracycline aptamer (*tet/metE*, Figure 4.5B). Engineering a chimera in which the length of the wild type *metE* P1 helix was preserved yielded a switch that is constitutively “OFF” under all tetracycline

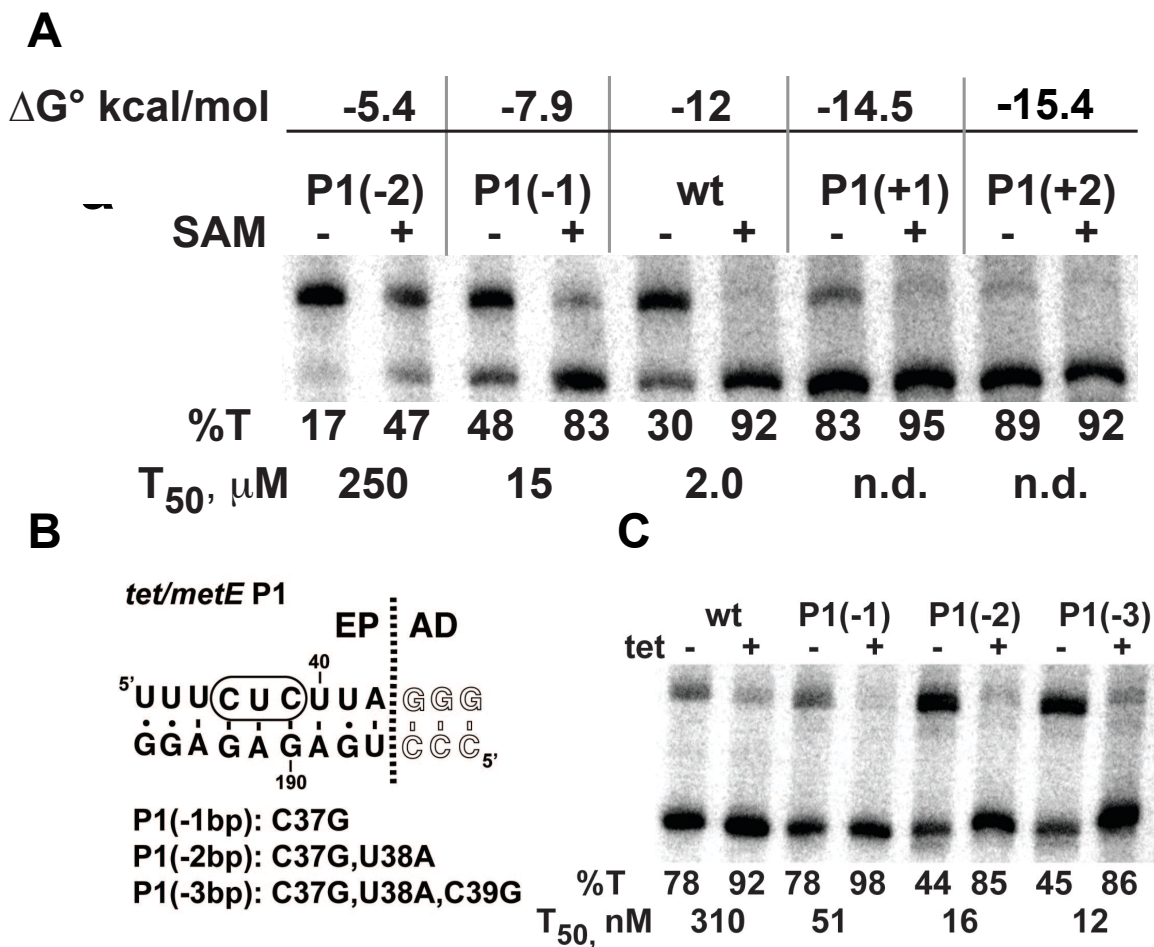


Figure 4.5 Modulation of P1 provides a general tuning mechanism

A) Results of P1 mutations of the *metE* riboswitch. Designation of the P1 mutations as -1 etc. is based on the diagram in figure 4.1 Free energy calculations were performed as described for the *lysC* riboswitch. B) The sequence of the P1 regulatory helix for a tetracycline aptamer (*tet*) fused to the *metE* expression platform. Mutations introduced into the 5'-side of the P1 helix to destabilize formation of the helix to favor formation of the competing antiterminator helix is shown below. (b) Transcription reactions at low and high tetracycline concentrations as a function of P1 helix length. Note that mutations that shorten the P1 helix by two or three base pairs enable a robust regulatory response. For each RNA, the T_{50} and percent termination at low and high effector concentrations are given as the average of three independent experiments.

concentrations. Based upon the data for *lysC* and *metE*, we created a set of weaker P1 helices and tested each for activity. While the *tet/metE*(P1-1) construct still displayed poor activity, but the P1-2 and P1-3 switches showed substantial switching (Figure 4.5C). Thus rational approaches can be taken to tune the regulatory responses of RNAs that utilize the proposed design strategy.

4.4- Chimeric riboswitches are functional in vivo.

To determine whether these RNAs are functional in cellular context, Joan Marcano in the lab examined the ability of a subset to regulate expression of a *gfpuv* reporter gene in *E. coli*. The parental reporter plasmid was constructed from pBR322 and contained the strong *rrnB* T₁T₂ terminators upstream of an insulated ptac promoter that would allow for constitutive initiation of transcription of the mRNA. Each riboswitch was placed immediately downstream of the promoter such the transcript started at the *metE*(+18) start site and the starting codon of *gfpuv* was placed at the start codon. Following transfection of *E. coli* BW25113(Δnep) with the reporter vector, cells were monitored for GFPuv fluorescence as a function of the appropriate effector titrated into the media. For these experiments, we focused on the 2-AP sensitive *xptC74U/metE* and theophylline sensitive *theo/metE* chimeras as many of the other aptamers sense ligands, such as lysine or SAM are present at concentrations that could potentially saturate the response signal in the absence of exogenous ligand. As negative controls, Joan introduced single point mutations into the ligand binding pocket of each aptamer (U51C for the *xpt* aptamer²⁰⁰ and U24A for the theophylline aptamer⁹⁸) that abrogate binding.

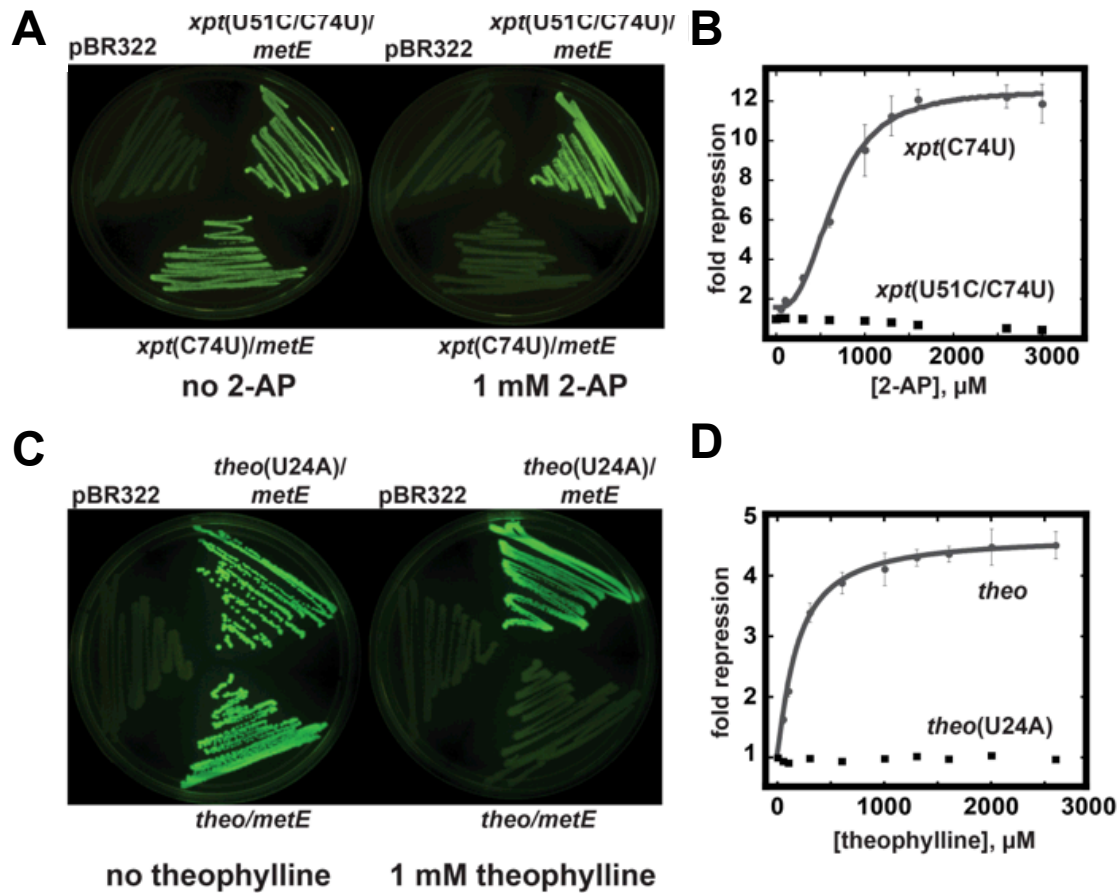


Figure 4.6 *In vivo* regulatory responses of chimeric RNAs

A) 2-aminopurine (2-AP) dependent regulation of *gfp* expression by the *xpt(C74U)/metE* chimera in *E. coli*. Cells transformed with either the control parental vector (pBR22) that does not contain the *gfp* gene, the *xpt(C74U)/metE* riboswitch in the 5'-leader of an mRNA encoding *gfpuv*, and a mutant of this riboswitch (U51C) that specifically abrogates 2-AP binding to the aptamer were plated on defined medium-agar in the absence (left) or presence (right) of 1 mM 2-AP. pBR322 serves as a control for the background fluorescence of *E. coli*, while the mutant is a control for potential non-specific theophylline dependence of *gfp* regulation that is not directly tied to binding of this molecule to the *xpt(C74U)* aptamer of the chimera. B) Quantification of the fluorescence of *E. coli* transformed with the *xpt(C74U)/metE* chimera (grey circles) or the mutant (U51C, black squares) grown in defined medium with increasing concentrations of 2-AP. C) Theophylline-dependent control of *gfp* expression in *E. coli*. Cells were grown under the same media conditions as in (a), but with a supplement of 1 mM theophylline in the right plate. A single point mutation in the theophylline aptamer (U24A) serves a control for potential theophylline-dependent effects on *gfp* expression that are not directly related to its binding to the aptamer. D) Quantification of the fluorescence of *E. coli* transformed with the *theo/metE* chimera (grey circles) or the point mutation thereof (black squares).

Joan demonstrated that these chimeric riboswitches display strong regulatory activity in *E. coli* (Figure 4.6), achieving full repression in the presence of 1 mM effector as compared to *E. coli* transfected with the pBR322 vector (Figure 4.6A and 4.6C). Furthermore, the point mutations completely abolish control *gfpuv* expression. Correction of the raw fluorescence data by subtracting the background fluorescence of cells containing the pBR322 vector yields data that can be fit to a modified two-state fitting equation (see Methods and Materials) to determine the EC₅₀ (concentration of ligand in the media required to elicit half-maximal repression of the reporter). For the *xpt* aptamer, an EC₅₀ of 730 ± 80 μM was measured, with a 12-fold dynamic range (Figure 4.6B); this EC₅₀ is ~25-fold higher than the T₅₀ measured *in vitro*. It must be noted that we utilized a strain of *E. coli* that was deficient in purine efflux (Δnep) and is still capable of purine degradation, and thus, the intracellular 2-AP concentration does not necessarily correlate with that in the medium, confounding any direct correlation between EC₅₀ and T₅₀. Similarly, the *theo/metE* riboswitch displayed a response of 190 ± 10 μM. While the value of EC₅₀ is higher than the observed T₅₀, it is consistent with the performance of other theophylline-dependent synthetic riboswitches^{190,201-203}.

4.5- Discussion.

It has been hypothesized that riboswitches are modular and their two functional domains, the ligand binding aptamer and the regulatory expression platform, can be “mixed-and-matched” to create RNA regulatory elements of novel activity. The modularity of riboswitch aptamer domains is clearly evidenced by the different regulatory contexts in which they are found. For example, the TPP riboswitch aptamer is

found in bacterial mRNA leader sequences to regulate either transcription or translation and in eukaryotes within introns or the 3'-UTR regulate alternative splicing or 3'-end processing⁴⁰. Despite these differences in how the riboswitch regulates expression, the bacterial and plant TPP aptamers are virtually identical^{62,63}. Clearly, nature is readily able to adapt various expression platforms onto a conserved aptamer. The Batey lab first demonstrated that this property can be exploited to create chimeric riboswitches using novel a novel aptamer-expression platform combination in a recent study in which a chimera between the *S. mutans folT* aptamer and the *B. subtilis metE* expression platform was used to demonstrate the regulatory activity of THF and guanine²⁰⁴. In the current work, we have greatly expanded upon this idea by examining a larger set of chimeras to determine the generality of this approach. Importantly, this work is the first demonstration that the two principal domains of riboswitches are modular in their functionality, which may have greatly facilitated their spread throughout the bacterial kingdom.

In the presented strategy for the design of chimeric riboswitches we sought to implement as effectively as possible engineering design principles that have been applied in the past towards other synthetic riboswitch systems²⁰⁵. *Scalability*, or the application of a sensory platform that can accommodate different ligand-sensory modules, has been demonstrated through the ability of three different expression platforms to respond to a variety of natural and artificial aptamers in a ligand-dependent fashion. *Portability*, the independence of the riboswitch to cell-specific machinery or regulatory mechanisms, has been established by virtue of expression platforms derived from *B. subtilis* to be able to direct *E. coli* RNAP both *in vitro* and *in vivo*. Since bacterial

RNAPs in general respond to intrinsic terminators of the type found in our expression platforms, we expect that these chimeras will be function in a broad spectrum of bacteria. *Utility*, or the coupling of the regulatory device to a functional output, is evidenced by their ability to control expression of *gfpuv in vivo*. In our design, our chimeras could be used to control expression of any conceivable downstream gene to elicit a desired output. *Composibility*, or a strategy to modularly couple aptamers and expression platforms in a way the preserves the activity of each, is clearly demonstrated through the ability to modulate the length of the P1 stem that connects the two units to achieve the desired ligand-dependent response. Importantly, these principles do not require selection or extensive screening of “communication modules” that are highly specific for the aptamer/expression platform pair. *Reliability*, or a means of standardizing the communication between the liganded state of the aptamer to the expression platform, is the greatest strength of this strategy for creating novel riboswitches. In it, we have preserved a set of P1 helix/expression platforms that have evolved to be highly effective in the cellular environment that is consistent between each aptamer/expression platform pair. The P1-antiterminator secondary structural switch is likely to be an extremely robust means of interdomain communication, as it is broadly utilized over almost classes of riboswitches found in every major clade of bacteria.

Another strength of this strategy for creation of artificial riboswitches is *flexibility* such that it can be extended to other forms of regulation or regulatory responses. For example, there are many expression platforms that control translation by exposing or occluding the ribosome binding site (RBS) using a similar secondary structural switch as

described above. Since translational regulation has been shown to be a viable means of generating cellular biosensors in other contexts, this may be a preferred regulatory mechanism for some applications. Furthermore, in synthetic biology, “ON” switches are considered to be extremely desirable. There are characterized riboswitches, such as those that respond to *S*-adenosylhomocysteine (SAH) that almost exclusively use expression platforms that turn genes on in response to their effector. One such platform that is part of a riboswitch that controls a purine efflux pump has been adapted for use in this strategy (Ceres and Batey, manuscript in preparation). Finally, expression platforms found in fungi and plants²⁰ could be adapted to this strategy to create novel regulatory devices for use in eukaryotic cells. The vast number and diversity of riboswitches and their associated expression platforms that have been identified provides a substantial reservoir of potential candidates for even more robust regulatory switches than the few characterized in this study.

4.6- Methods and Materials

Synthesis of DNA templates for transcription

The DNA templates used for single turnover transcription assays were synthesized by PCR amplification from *B. subtilis* genomic DNA in the case of the wild type sequences. These sequences were placed immediately downstream of the T7A1 promoter using recombinant PCR. Chimeric riboswitches were constructed in two different pieces. The first piece consists of the T7A1 promoter, the 5'-side of the expression platform of choice (*metE*, *yitJ* or *lysC*), and the aptamer domain. The second piece was composed of the 3'-side of the expression platform. Recombinant PCR was used to fuse these two portions. Mutant templates were obtained by altering the desired position/s in the DNA by standard site directed mutagenesis. The sequences of all DNAs were verified prior to use in transcription assays.

In vitro transcription assays

DNA templates containing a chimeric riboswitch were transcribed as described²⁰⁴. Briefly, 100 ng of DNA were incubated at 37 °C for 15 minutes in 25 µl of 2X transcription buffer (140 mM Tris-HCl, pH 8.0, 140 mM NaCl, 0.2 mM EDTA, 28 mM β-mercaptoethanol and 70 mg/mL BSA), 5 µl of 25 mM MgCl₂, 1 mCi of α-³²P-ATP and 0.5 units of *E. coli* RNA polymerase 70 holoenzyme (Epicentre Biotechnologies) to a

final volume of 35 ml. The reactions were initiated with the addition of 15 ml of NTP mix (NTPs in equimolar ratios of 165 mM), 0.2 mg/ml heparin and the desired concentration of ligand. After incubation at 37 °C for 15 minutes the reactions were quenched with equal volume of 8 M urea and 2 minutes incubation at 65 °C. The species were separated by denaturing PAGE, dried and exposed to a phosphorimager screen. Quantitation of radioactive counts in each band was performed with ImageQuant software (Molecular Biosystems) and data fit to a two-state model with nonlinear least squares analysis.

In vivo assay

A reporter plasmid for testing riboswitch function in *E. coli* with a *gfpuv* reporter gene was constructed using pBR322 as the parental vector using standard molecular biological techniques. The sequence of the chimeric riboswitch insert for the *xpt*(C74U)/*metE* and *theo/metE* reporters are identical to those used for *in vitro* transcription assays. The resultant vectors were transformed into *E. coli* strain BW25113(Δnep). Single colonies were picked and grown overnight in 3 ml of CSB media + 100 µg/ml ampicillin. This saturated culture was used to inoculate 100 ml of fresh media, and allowed to grow to early exponential phase ($OD_{600} = 0.1-0.5$). At this point the culture was aliquoted to 3 ml in triplicate and the ligand was added to the media at the concentrations indicated in the titration. The cells were allowed to grow for 6 hours at 37 °C. At this point 300 µl of the cultures were used to measure the OD_{600} of their fluorescent intensity in a plate reader (Tecan). Fluorescent measurements were taken at an excitation wavelength of 395 nm and the average fluorescence was taken from 513-515 nm, where the max emission for GFP was observed. Optical density normalized fluorescent values were plotted against the ligand concentration where a curve was generated and fitted to a two state binding equation to extract the EC_{50} values. The background fluorescence was obtained by performing a ligand titration into cells carrying the parental plasmid pBR322 and calculating the OD_{600} normalized fluorescence. An average of the background was subtracted to the values of the cultures containing the plasmids with the riboswitch constructs.

Chapter 5: Conclusions and future directions

5.1 Structural similarities between lysine recognition by RNA and proteins

The RNA world hypothesis posits that RNAs formed the basis of a primitive protometabolism which constituted the first life on earth, pre-dating the so called last common ancestor (LUCA) that gave rise to all modern lineages²⁰⁶⁻²⁰⁹. While no direct fossil records remain of this alien world, many lines of evidence support the probability of its existence. For example, RNA is the only macromolecule that can carry the inheritable information *and* enact phenotypic changes by performing catalytic or regulatory functions in a specific fashion. It is intimately associated with many of the processes that are central to information transfer, playing critical roles in the catalysis of protein synthesis, tRNA processing and mRNA splicing. Perhaps most importantly RNA

Other arguments can be made on the basis that RNA based metabolism is highly probable due to the small amount of RNA required to encompass a large variety of potentially functional sequence motifs²¹⁰, and the high probability that these motifs would exist in a relatively small collection of random sequences. Indeed *in vitro* selection experiments have recently demonstrated that an RNA of only 5 nucleotides can catalyze an amino acylation reaction reminiscent of the catalytic functions performed by modern tRNA synthetases²¹¹. Most extant RNAs however, provide a somewhat vague picture of how the RNA world was constructed as proteins have supplanted the many of the metabolic functions of their RNA precursors, particularly in terms of catalysis²¹². On the other hand many riboswitches bind to fundamentally important metabolites that likely arose early in the evolution of biological systems^{16,213},

suggesting that these RNAs may be rare survivors of this ancient time. While at present there are no clear means by which to directly test such speculation, riboswitch aptamers nonetheless provide a unique opportunity to compare and contrast the principles of molecular recognition by RNA and the proteins that have supplanted them.

The lysine riboswitch provides the first example of a lysine binding RNA aptamer, as it make a particularly difficult target to specifically select by *in vitro* methods due to it's positively charged character. This aptamer makes direct contacts with all of the available functional groups of the amino acid, however data presented in chapter 3 demonstrates the emphasis placed on recognition of the side chain: a total of four electrostatic interactions between the G77 backbone and the ϵ -amine appear to provide a relatively strong set of non-covalent bonds. This feature of the RNA ensures a high level of amino acid specificity, despite the relatively limited chemical diversity of RNA nucleotides that can contribute to recognition.

Another biological process that requires exquisite specificity is the selection of amino acids by their cognate tRNA synthetases (RS), which serve as the molecular machines that couple amino acids to their appropriate tRNA substrates²¹⁴. Comparison of the binding pockets of the *E. coli* Lys-RS protein²¹⁵ and the lysine riboswitch (Figure 5.1, PDB ID: 1BBU) shows that both macromolecules appear to place emphasis on recognition of the side chain while having more sparse contacts with the main chain moieties. The requirement for high fidelity translation presumably places a premium on avoiding promiscuous amino acid binding by avoiding strong recognition of the main chain atoms similar to what I have described for the lysine riboswitch. In support of this hypothesis both the lysine riboswitch and lysyl-RS enzymes have been found to bind

main chain analogs more tightly than derivatives that alter the position or charge of the ϵ -amine²¹⁶ (Table 5.1). Both structures are also implicated as targets of AEC as they are unable to strongly discriminate this compound from lysine^{216,217}.

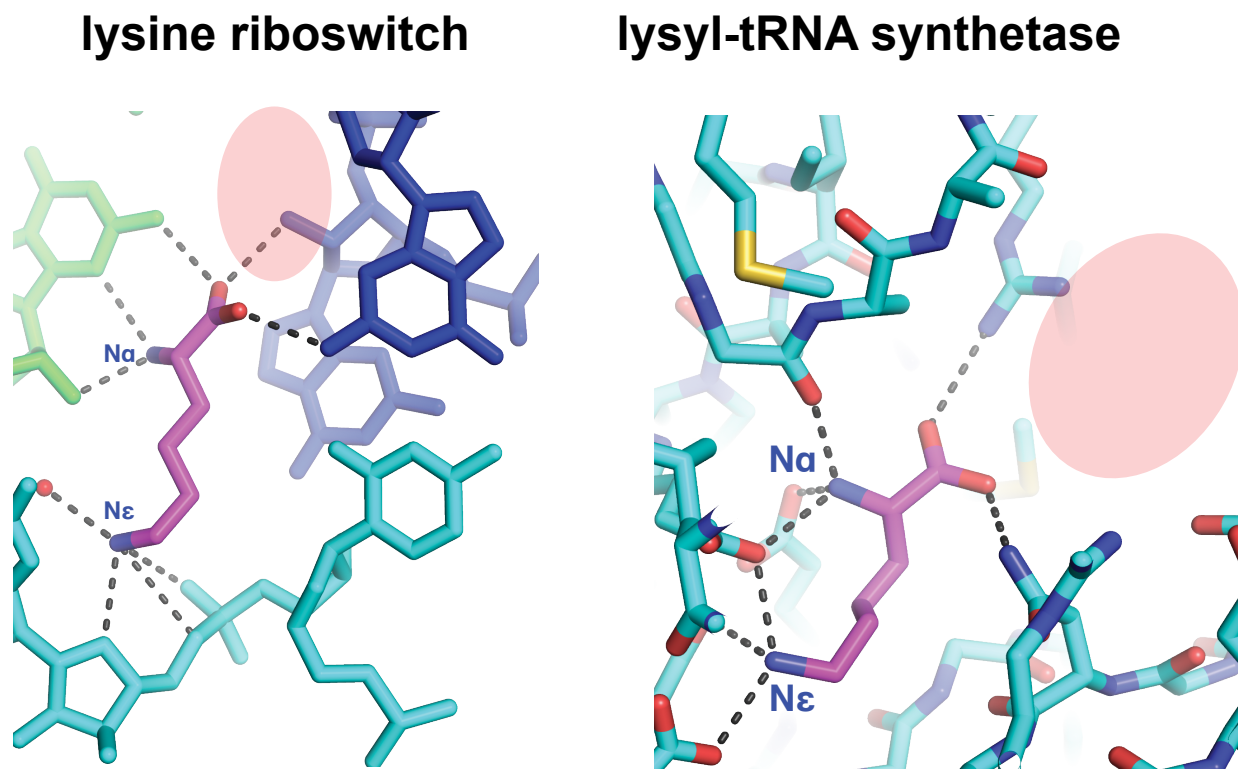


Figure 5.1 Comparison of lysine binding pockets the lysine riboswitch and lysyl-tRNA synthetase

Side chain and main chain amines are labeled for reference. Red ovals indicate cavities adjacent to the carboxyl moiety of lysine that are common to these structures.

In addition to the requirements for translational fidelity, weak recognition of the main chain atoms by the lysRS enzyme may also be a consequence of the need to maintain a large cleft for ATP binding near the carboxyl terminus (Red oval Figure 5.1). ATP serves as the donor for adenylation of the amino acid at the carboxyl terminus, activating it for attachment to the 3' end of the cognate tRNA. Coincidentally the lysine riboswitch also has a cavity near this end of the ligand that can still accommodate an additional glycyl moiety. This observation provokes interesting questions about the

Table 5.1 Molecular recognition by the lysine riboswitch and lysRS enzymes

	Compound	Lysine riboswitch K_D (μ M)	^a lysRS1 K_i (μ M)	^a lysRS2 K_i (μ M)
Main chain	L-Lysine	53 $\pm$ 6	*34 $\pm$ 3	*2.6 $\pm$ 0.2
	Lysinamide	90 $\pm$ 30	2120 $\pm$ 450	17 $\pm$ 2
	L-lysine methyl ester	110 $\pm$ 30	478 $\pm$ 100	74 $\pm$ 7
	L-lysine ethyl ester	250 $\pm$ 30	303 $\pm$ 45	55 $\pm$ 6
	<i>D-Lysine</i>	n.d. in this work	6900 $\pm$ 2500	12000 $\pm$ 1400
Side chain	S-(2-Aminoethyl)-L-cysteine	**~830	1140 $\pm$ 230	3.9 $\pm$ 0.4
	Ornithine	not observed	8800 $\pm$ 1300	6300 $\pm$ 600
	Arginine	not observed	5060 $\pm$ 860	64000 $\pm$ 5000

^aValues from Levengood *et. al.* **JBC** (2004)

*Indicates K_m as opposed to K_D

*Value taken from Ataide *et. al.* ACS Chem. Biol. (2007). For reference the K_D reported for lysine in this paper was 360 μ M.

origins of this aptamer, and whether ancestral sequences not only bound lysine, but utilized it to catalyze amino acid condensation or activation; reactions that were subsequently elaborated on with the emergence of proteins. Indeed many riboswitches bind molecules that serve as cofactors in extant metabolism suggesting they may have once harnessed the chemical potential of their ligands to coordinate more complex metabolic reactions in the RNA world than those that have persisted in existing ribozymes^{218,219}. The *glmS* riboswitch may actually represent such a case of that has persisted into modern biology as it utilizes glucosamine-6-phosphate as a cofactor for mediating the catalytic cleavage to regulate gene expression²¹⁸.

On the other hand, the sparse phylogenetic distribution of many riboswitch classes, including the lysine riboswitch³⁷ (Figure 1.1) poses a significant challenge to

the view that riboswitches have descended from ancient origins, and could reflect a number of alternative evolutionary scenarios. For example, though no direct evidence has been yet been put forth, the phylogenetic distribution of many riboswitches could reflect lateral transfer events that lead to acquisition of these elements in a few distinct bacterial phyla. Then again, it is also possible that the ancient regulatory functions performed by these RNAs were replaced by protein based mechanisms, though evidence for this scenario would likely be lost over the course of long periods of evolution. Future bioinformatics efforts will hopefully lend deeper insight into these issues, and broaden our understanding of the true origins of these remarkable RNAs.

5.2 Riboswitches in the cellular milieu: where do we go from here?

While structural and biochemical characterization of riboswitch–ligand interactions have illuminated many aspects of their regulatory mechanisms, these data cannot always be completely reconciled with their biological activity. This is exemplified in a recent study of the variability in the ligand responsiveness of 11 different SAM responsive transcriptional units in *B. subtilis* genome¹⁰⁸. These riboswitches were demonstrated to have a ~250 fold range in KD and T₅₀ *in vitro*. Additionally, *in vivo* Northern blotting, qRT-PCR, and lacZ reporter analyses revealed a large degree of variability in the degree of the termination response upon addition of methionine: this ranged from a 1.2-fold increase in termination for the *metK* gene in the presence of methionine, to a 340-fold increase for the *metE* gene as measured by qRT-PCR. The function of the gene products is thought to be tied to this disparity. Genes involved directly in methionine biosynthesis experienced the tightest regulation, the greatest level

of induction, and the longest delays before induction could be detected. Meanwhile riboswitches controlling methionine transport and genes of unknown function displayed lower levels of repression during growth in methionine supplemented media, and lower magnitudes of induction during methionine starvation. This study suggests that there are structural differences in the aptamer domain or expression platform of these SAM riboswitches may account for these disparities, motivating the need to correlate these observations to available structural and sequence based data. Furthermore, differences in the time of transcription (i.e. transcriptional pausing) at these different genetic loci may tune the response, requiring further characterization of this phenomenon for each transcript.

Interestingly, not all of these riboswitches behaved as expected. The *cysH* gene does not experience induction upon methionine starvation¹⁰⁸, which may be explained by previous reports that this operon is regulated primarily at the DNA level by a transcriptional repressor protein that responds to levels of the cysteine precursor O-acetyl-L-serine 70. The *metK* gene on the other hand experiences a transient rise in read through product at 0.5 h, but falls back to basal levels after 1 h¹⁰⁸. This expression pattern suggests that this transcript may also be regulated at either the translational level or by the effects of mRNA degradation. A consequence of overlapping post-transcriptional regulation may therefore lead to differences in riboswitch response efficiency, further highlighting the need to correlate biochemical data with biological activity.

Another motivation for studying ligand interactions in their cellular context is to determine the potential of these RNAs as novel antimicrobial targets. Ribosomal RNA

sets a precedence for RNA based strategies, as its known to be the target of aminoglycoside antibiotics, and many riboswitches have already been linked to antibiotic effects of naturally occurring compounds. For example, the lysine riboswitch has long been recognized as a mutational hotspot in microbes resistant to the lysine analogue AEC^{123,220}. A recent study implicated role of the lysine riboswitch mediating the toxicity of this and other lysine analogues in *B. subtilis*⁸⁹. However, shortly after this study was published, it was discovered that the primary target of these compounds in *E. coli* was one of the two lysyl-tRNA synthetase variants (LysRS) in this organism and the toxic effects are due to incorporation into nascent polypeptides during translation²¹⁷. Mutations to the lysine riboswitch cause inefficient repression of lysine biosynthesis genes, thus allowing lysine pools to become elevated and effectively compete for the charging of LysRS²¹⁷. A similar study demonstrated that the TPP riboswitch serves as a mutational hotspot for relieving the toxic effects of the TPP analogue pyrithiamine pyrophosphate (PTPP) in *B. subtilis*, though many of the PTPP resistant *E. coli* mutants that were sequenced apparently acquired resistance by some other mechanism²²¹. However, the toxic effects of roseoflavin, an FMN analogue, to *B. subtilis* are at least in part due to direct targeting of the riboswitch upstream of the ribD operon by the antimicrobial²²². The homologous operon in *Streptomyces davawensis*- an organism that naturally produces this antibacterial agent-is also responsive to roseoflavin²²² leaving the mechanism of this bacteria's natural resistance to this antimicrobial compound unclear and highlighting the need to understand the biological roles of riboswitches in their cellular context.

With a wealth of structural information available (currently, at least 14 individual aptamer–ligand complexes have been solved by X-ray crystallography), a significant challenge remains to correlate these data with in vitro and in vivo studies that take into account the temporal aspects of riboswitch synthesis and function. In particular, new approaches need to be developed that can readily monitor cotranscriptional riboswitch folding with an eye towards its relationship to rates of ligand binding, the role of transcriptional pausing in the expression platform, and secondary structural rearrangements and their relationship to efficient genetic regulation. Determination of response kinetics and regulatory capacity (i.e. fold induction or repression) in vivo may also reveal factors that overlap with riboswitch regulation to further tune expression; concepts that will need to be addressed to determine efficacy of targeting riboswitches for therapeutic purposes. An overarching goal of these studies will be to provide a clear link between structural and functional studies that takes into consideration the complicated folding process of a riboswitch in the cell.

Works Cited

1. Sharp, P. A. The Centrality of RNA. *Cell* **136**, 577–580 (2009).
2. Eddy, S. R. Non-coding RNA genes and the modern RNA world. *Nat Rev Genet* **2**, 919–929 (2001).
3. Bartel, D. P. & Chen, C.-Z. Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat Rev Genet* **5**, 396–400 (2004).
4. Bartel, D. P. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**, 281–297 (2004).
5. Carthew, R. W. & Sontheimer, E. J. Origins and Mechanisms of miRNAs and siRNAs. *Cell* **136**, 642–655 (2009).
6. Friedman, R. C., Farh, K. K.-H., Burge, C. B. & Bartel, D. P. Most mammalian mRNAs are conserved targets of microRNAs. *Genome Research* **19**, 92–105 (2009).
7. Mercer, T., Dinger, M. & Mattick, J. Long non-coding RNAs: insights into functions. *Nat Rev Genet* (2009).doi:10.1038/nrg2521
8. Mattick, J. S. A new paradigm for developmental biology. *Journal of Experimental Biology* **210**, 1526–1547 (2007).
9. Taft, R. J., Pheasant, M. & Mattick, J. S. The relationship between non-protein-coding DNA and eukaryotic complexity. *Bioessays* **29**, 288–299 (2007).
10. Gottesman, S. Micros for microbes: non-coding regulatory RNAs in bacteria. *Trends in Genetics* **21**, 399–404 (2005).
11. Waters, L. S. & Storz, G. Regulatory RNAs in bacteria. *Cell* **136**, 615–628 (2009).
12. Wiedenheft, B., Sternberg, S. H. & Doudna, J. A. RNA-guided genetic silencing systems in bacteria and archaea. *Nature* **482**, 331–338 (2012).
13. Marraffini, L. A. & Sontheimer, E. J. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet* **11**, 181–190 (2010).
14. Timmons, L. & Fire, A. Specific interference by ingested dsRNA. *Nature* **395**, 854 (1998).
15. Gold, L. *et al.* Aptamers and the RNA world, past and present. *Cold Spring Harbor perspectives in biology* **4**, (2012).
16. Winkler, W. C. & Breaker, R. R. Regulation of bacterial gene expression by riboswitches. *Annu Rev Microbiol* **59**, 487–517 (2005).
17. Kertsburg, I. A. & Winkler, W. C. Genetic Control by cis-Acting Regulatory RNAs in *Bacillus subtilis*: General Principles and Prospects for Discovery. *Cold Spring Harb Symp Quant Biol* **71**, 239–249 (2006).
18. Stoddard, C. D. & Batey, R. T. Mix-and-match riboswitches. *ACS Chem Biol* **1**, 751–754 (2006).
19. Mandal, M., Boese, B., Barrick, J. E., Winkler, W. C. & Breaker, R. R. Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**, 577–586 (2003).

20. Wachter, A. *et al.* Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell* **19**, 3437–3450 (2007).
21. Mandal, M. & Breaker, R. R. Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* **11**, 29–35 (2004).
22. Sudarsan, N., Wickiser, J. K., Nakamura, S., Ebert, M. S. & Breaker, R. R. An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes Dev* **17**, 2688–2697 (2003).
23. Rodionov, D. A., Vitreschak, A. G., Mironov, A. A. & Gelfand, M. S. Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch? *Nucleic Acids Res* **31**, 6748–6757 (2003).
24. Mandal, M. *et al.* A glycine-dependent riboswitch that uses cooperative binding to control gene expression. *Science* **306**, 275–279 (2004).
25. Ames, T. D. & Breaker, R. R. Bacterial aptamers that selectively bind glutamine. *RNA Biol* **8**, 82–89 (2011).
26. Nahvi, A. *et al.* Genetic control by a metabolite binding mRNA. *Chem Biol* **9**, 1043 (2002).
27. Winkler, W., Nahvi, A. & Breaker, R. R. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature* **419**, 952–956 (2002).
28. Mironov, A. S. *et al.* Sensing small molecules by nascent RNA: a mechanism to control transcription in bacteria. *Cell* **111**, 747–756 (2002).
29. Ames, T. D., Rodionov, D. A., Weinberg, Z. & Breaker, R. R. A eubacterial riboswitch class that senses the coenzyme tetrahydrofolate. *Chem Biol* **17**, 681–685 (2010).
30. Winkler, W. C., Nahvi, A., Roth, A., Collins, J. A. & Breaker, R. R. Control of gene expression by a natural metabolite-responsive ribozyme. *Nature* **428**, 281–286 (2004).
31. Dann, C. E. *et al.* Structure and mechanism of a metal-sensing regulatory RNA. *Cell* **130**, 878–892 (2007).
32. Baker, J. L. *et al.* Widespread genetic switches and toxicity resistance proteins for fluoride. *Science's STKE* **335**, 233 (2012).
33. Sudarsan, N. *et al.* Riboswitches in eubacteria sense the second messenger cyclic di-GMP. *Science* **321**, 411–413 (2008).
34. Weinberg, Z. *et al.* Identification of 22 candidate structured RNAs in bacteria using the CMfinder comparative genomics pipeline. *Nucleic Acids Res* **35**, 4809–4819 (2007).
35. Weinberg, Z. *et al.* Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol* **11**, R31 (2010).
36. Dambach, M. D. & Winkler, W. C. Expanding roles for metabolite-sensing regulatory RNAs. *Curr Opin Microbiol* **12**, 161–169 (2009).
37. Barrick, J. E. & Breaker, R. R. The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol* **8**, R239 (2007).

38. Collins, J. A., Irnov, I., Baker, S. & Winkler, W. C. Mechanism of mRNA destabilization by the glmS ribozyme. *Genes Dev* **21**, 3356–3368 (2007).
39. McCarthy, T. J. *et al.* Ligand requirements for glmS ribozyme self-cleavage. *Chem Biol* **12**, 1221–1226 (2005).
40. Cheah, M. T., Wachter, A., Sudarsan, N. & Breaker, R. R. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**, 497–500 (2007).
41. Lee, E. R., Baker, J. L., Weinberg, Z., Sudarsan, N. & Breaker, R. R. An Allosteric Self-Splicing Ribozyme Triggered by a Bacterial Second Messenger. *Science* **329**, 845–848 (2010).
42. Tucker, B. J. & Breaker, R. R. Riboswitches as versatile gene control elements. *Curr Opin Struct Biol* **15**, 342–348 (2005).
43. Sudarsan, N. *et al.* Tandem riboswitch architectures exhibit complex gene control functions. *Science* **314**, 300–304 (2006).
44. Fox, K. A. *et al.* Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in *Enterococcus faecalis*. *Proceedings of the National Academy of Sciences* **106**, 4435–4440 (2009).
45. Ray, P. S. *et al.* A stress-responsive RNA switch regulates VEGFA expression. *Nature* **457**, 915–919 (2009).
46. Blount, K. F. & Breaker, R. R. Riboswitches as antibacterial drug targets. *Nat Biotechnol* **24**, 1558–1564 (2006).
47. Moore, P. STRUCTURAL MOTIFS IN RNA. *Annu Rev Biochem* (1999).
48. Lescoute, A., Leontis, N. B., Massire, C. & Westhof, E. Recurrent structural RNA motifs, Isostericity Matrices and sequence alignments. *Nucleic Acids Res* **33**, 2395–2409 (2005).
49. Lescoute, A. & Westhof, E. The interaction networks of structured RNAs. *Nucleic Acids Res* **34**, 6587–6604 (2006).
50. Klein, D. J., Schmeing, T. M., Moore, P. B. & Steitz, T. A. The kink-turn: a new RNA secondary structure motif. *EMBO J* **20**, 4214–4221 (2001).
51. Leontis, N. B. & Westhof, E. The 5S rRNA loop E: chemical probing and phylogenetic data versus crystal structure. *RNA* **4**, 1134–1153 (1998).
52. Jaeger, L., Michel, F. & Westhof, E. Involvement of a GNRA tetraloop in long-range RNA tertiary interactions. *J Mol Biol* **236**, 1271–1276 (1994).
53. Tuerk, C. *et al.* CUUCGG hairpins: extraordinarily stable RNA secondary structures associated with various biochemical processes. *Proc Natl Acad Sci U S A* **85**, 1364–1368 (1988).
54. Nissen, P., Ippolito, J. A., Ban, N., Moore, P. B. & Steitz, T. A. RNA tertiary interactions in the large ribosomal subunit: the A-minor motif. *Proc Natl Acad Sci USA* **98**, 4899–4903 (2001).
55. Cate, J. H. *et al.* Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science* **273**, 1678–1685 (1996).
56. Costa, M. & Michel, F. Frequent use of the same tertiary motif by self-folding RNAs. *EMBO J* (1995).
57. Cromie, M. J., Shi, Y., Latifi, T. & Groisman, E. A. An RNA sensor for intracellular Mg(2+). *Cell* **125**, 71–84 (2006).

58. Klein, D. J. & Ferré-D'Amaré, A. R. Structural basis of glmS ribozyme activation by glucosamine-6-phosphate. *Science* **313**, 1752–1756 (2006).
59. Thore, S., Leibundgut, M. & Ban, N. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. *Science* **312**, 1208–1211 (2006).
60. Batey, R. T., Gilbert, S. D. & Montange, R. K. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature* **432**, 411–415 (2004).
61. Martick, M. & Scott, W. G. Tertiary contacts distant from the active site prime a ribozyme for catalysis. *Cell* **126**, 309–320 (2006).
62. Thore, S., Frick, C. & Ban, N. Structural Basis of Thiamine Pyrophosphate Analogues Binding to the Eukaryotic Riboswitch. *J Am Chem Soc* (2008).doi:10.1021/ja801708e
63. Serganov, A., Polonskaia, A., Phan, A. T., Breaker, R. R. & Patel, D. J. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* **441**, 1167–1171 (2006).
64. Lescoute, A. & Westhof, E. Topology of three-way junctions in folded RNAs. *RNA* **12**, 83–93 (2006).
65. la Peña, de, M., Dufour, D. & Gallego, J. Three-way RNA junctions with remote tertiary contacts: a recurrent and highly versatile fold. *RNA* **15**, 1949–1964 (2009).
66. Brierley, I., Pennell, S. & Gilbert, R. J. C. Viral RNA pseudoknots: versatile motifs in gene expression and replication. *Nat Rev Micro* **5**, 598–610 (2007).
67. Theimer, C. A., Blois, C. A. & Feigon, J. Structure of the human telomerase RNA pseudoknot reveals conserved tertiary interactions essential for function. *Mol Cell* **17**, 671–682 (2005).
68. Theimer, C. A. & Feigon, J. Structure and function of telomerase RNA. *Curr Opin Struct Biol* **16**, 307–318 (2006).
69. Gilbert, S. D., Rambo, R. P., Van Tyne, D. & Batey, R. T. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. *Nat Struct Mol Biol* **15**, 177–182 (2008).
70. Kang, M., Peterson, R. & Feigon, J. Structural Insights into Riboswitch Control of the Biosynthesis of Queuosine, a Modified Nucleotide Found in the Anticodon of tRNA. *Mol Cell* (2009).doi:10.1016/j.molcel.2009.02.019
71. Klein, D., Edwards, T. & Ferré-D'Amaré, A. Cocystal structure of a class I preQ(1) riboswitch reveals a pseudoknot recognizing an essential hypermodified nucleobase. *Nat Struct Mol Biol* (2009).doi:10.1038/nsmb.1563
72. Krasilnikov, A. S. On the occurrence of the T-loop RNA folding motif in large RNA molecules. *RNA* **9**, 640–643 (2003).
73. Adams, P. L., Stahley, M. R., Kosek, A. B., Wang, J. & Strobel, S. A. Crystal structure of a self-splicing group I intron with both exons. *Nature* **430**, 45–50 (2004).
74. Kwon, M. & Strobel, S. A. Chemical basis of glycine riboswitch cooperativity. *RNA* **14**, 25–34 (2008).

75. Montange, R. K. & Batey, R. T. Structure of the S-adenosylmethionine riboswitch regulatory mRNA element. *Nature* **441**, 1172–1175 (2006).
76. Pan, T. & Sosnick, T. R. Intermediates and kinetic traps in the folding of a large ribozyme revealed by circular dichroism and UV absorbance spectroscopies and catalytic activity. *Nat Struct Biol* **4**, 931–938 (1997).
77. Treiber, D. K. & Williamson, J. R. Beyond kinetic traps in RNA folding. *Curr Opin Struct Biol* **11**, 309–314 (2001).
78. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. RNA SHAPE chemistry reveals nonhierarchical interactions dominate equilibrium structural transitions in tRNA(Asp) transcripts. *J Am Chem Soc* **127**, 4659–4667 (2005).
79. Wilkinson, K. A., Merino, E. J. & Weeks, K. M. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. *Nat Protoc* **1**, 1610–1616 (2006).
80. Stoddard, C. D., Gilbert, S. D. & Batey, R. T. Ligand-dependent folding of the three-way junction in the purine riboswitch. *RNA* **14**, 675–684 (2008).
81. Merino, E. J., Wilkinson, K. A., Coughlan, J. L. & Weeks, K. M. RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). *J Am Chem Soc* **127**, 4223–4231 (2005).
82. Brion, P. & Westhof, E. Hierarchy and dynamics of RNA folding. *Annu Rev Bioph Biom* **26**, 113–137 (1997).
83. Lemay, J.-F., Penedo, J. C., Tremblay, R., Lilley, D. M. J. & Lafontaine, D. A. Folding of the adenine riboswitch. *Chem Biol* **13**, 857–868 (2006).
84. Ryals, J., Little, R. & Bremer, H. Temperature dependence of RNA synthesis parameters in *Escherichia coli*. *J Bacteriol* **151**, 879–887 (1982).
85. King, G. M., Carter, A. R., Churnside, A. B., Eberle, L. S. & Perkins, T. T. Ultrastable atomic force microscopy: atomic-scale stability and registration in ambient conditions. *Nano Lett* **9**, 1451–1456 (2009).
86. Li, P., Viereggs, J. & Tinoco, I., Jr How RNA Unfolds and Refolds. *Annu Rev Biochem* (2008).
87. Greenleaf, W. J., Frieda, K. L., Foster, D. A. N., Woodside, M. T. & Block, S. M. Direct Observation of Hierarchical Folding in Single Riboswitch Aptamers. *Science* **319**, 630–633 (2008).
88. Pan, T. & Sosnick, T. RNA folding during transcription. *Annu Rev Bioph Biom* **35**, 161–175 (2006).
89. Blount, K. F., Wang, J. X., Lim, J., Sudarsan, N. & Breaker, R. R. Antibacterial lysine analogs that target lysine riboswitches. *Nat Chem Biol* **3**, 44–49 (2007).
90. Kim, J. N. & Breaker, R. R. Purine sensing by riboswitches. *Biol. Cell* **100**, 1–11 (2008).
91. Kim, J. N., Roth, A. & Breaker, R. R. Guanine riboswitch variants from *Mesoplasma florum* selectively recognize 2'-deoxyguanosine. *Proc Natl Acad Sci USA* **104**, 16092–16097 (2007).

92. Serganov, A. *et al.* Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. *Chem Biol* **11**, 1729–1741 (2004).
93. Serganov, A., Huang, L. & Patel, D. J. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. *Nature* **458**, 233–237 (2009).
94. Edwards, A. L. & Batey, R. T. A structural basis for the recognition of 2'-deoxyguanosine by the purine riboswitch. *J Mol Biol* **385**, 938–948 (2009).
95. Serganov, A., Huang, L. & Patel, D. J. Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature* **455**, 1263–1267 (2008).
96. Edwards, T. E., Klein, D. J. & Ferré-D'Amaré, A. R. Riboswitches: small-molecule recognition by gene regulatory RNAs. *Curr Opin Struct Biol* **17**, 273–279 (2007).
97. Jenison, R. D., Gill, S. C., Pardi, A. & Polisky, B. High-resolution molecular discrimination by RNA. *Science* **263**, 1425–1429 (1994).
98. Zimmermann, G. R., Jenison, R. D., Wick, C. L., Simorre, J. P. & Pardi, A. Interlocking structural motifs mediate molecular discrimination by a theophylline-binding RNA. *Nat Struct Biol* **4**, 644–649 (1997).
99. Williamson, J. R. Induced fit in RNA-protein recognition. *Nat Struct Biol* **7**, 834–837 (2000).
100. Noeske, J., Schwalbe, H. & Wohnert, J. Metal-ion binding and metal-ion induced folding of the adenine-sensing riboswitch aptamer domain. *Nucleic Acids Res* **35**, 5262–5273 (2007).
101. Noeske, J. *et al.* Interplay of “induced fit” and preorganization in the ligand induced folding of the aptamer domain of the guanine binding riboswitch. *Nucleic Acids Res* **35**, 572–583 (2007).
102. Ottink, O. M. *et al.* Ligand-induced folding of the guanine-sensing riboswitch is controlled by a combined predetermined induced fit mechanism. *RNA* **13**, 2202–2212 (2007).
103. Noeske, J. *et al.* An intermolecular base triple as the basis of ligand specificity and affinity in the guanine- and adenine-sensing riboswitch RNAs. *Proc Natl Acad Sci USA* **102**, 1372–1377 (2005).
104. Lang, K., Rieder, R. & Micura, R. Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach. *Nucleic Acids Res* **35**, 5370–5378 (2007).
105. Baird, N. J. & Ferré-D'Amaré, A. R. Idiosyncratically tuned switching behavior of riboswitch aptamer domains revealed by comparative small-angle X-ray scattering analysis. *RNA* **16**, 598–609 (2010).
106. Bailor, M. H., Sun, X. & Al-Hashimi, H. M. Topology links RNA secondary structure with global conformation, dynamics, and adaptation. *Science* **327**, 202–206 (2010).
107. Blouin, S. & Lafontaine, D. A. A loop loop interaction and a K-turn motif located in the lysine aptamer domain are important for the riboswitch gene regulation control. *RNA* **13**, 1256–1267 (2007).

108. Tomsic, J., McDaniel, B. A., Grundy, F. J. & Henkin, T. M. Natural variability in S-adenosylmethionine (SAM)-dependent riboswitches: S-box elements in bacillus subtilis exhibit differential sensitivity to SAM In vivo and in vitro. *J Bacteriol* **190**, 823–833 (2008).
109. Wickiser, J. K., Winkler, W. C., Breaker, R. R. & Crothers, D. M. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* **18**, 49–60 (2005).
110. Epshtein, V. & Mironov, A. The riboswitch-mediated control of sulfur metabolism in bacteria. *Proceedings of the ...* (2003).
111. Gilbert, S. D., Stoddard, C. D., Wise, S. J. & Batey, R. T. Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. *J Mol Biol* **359**, 754–768 (2006).
112. Wickiser, J. K., Cheah, M. T., Breaker, R. R. & Crothers, D. M. The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry* **44**, 13404–13414 (2005).
113. Wong, T. N., Sosnick, T. R. & Pan, T. Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc Natl Acad Sci USA* **104**, 17995–18000 (2007).
114. Xayaphoummine, A., Viasnoff, V., Harlepp, S. & Isambert, H. Encoding folding paths of RNA switches. *Nucleic Acids Res* **35**, 614–622 (2007).
115. Crothers, D., Cole, P., Hilbers, C. & Shulman, R. The molecular mechanism of thermal unfolding of Escherichia coli formylmethionine transfer RNA. *J Mol Biol* **87**, 63 (1974).
116. Mulhbach, J. & Lafontaine, D. A. Ligand recognition determinants of guanine riboswitches. *Nucleic Acids Res* **35**, 5568–5580 (2007).
117. Breaker, R. R. Riboswitches and the RNA world. *Cold Spring Harbor perspectives in biology* **4**, (2012).
118. Serganov, A. & Patel, D. J. Amino acid recognition and gene regulation by riboswitches. *Biochim Biophys Acta* **1789**, 592–611 (2009).
119. Domigan, L. J. *et al.* Characterisation of dihydrodipicolinate synthase (DHDPS) from Bacillus anthracis. *Biochimica et Biophysica Acta (BBA) - Proteins & Proteomics* **1794**, 1510–1516 (2009).
120. Cox, R., Sutherland, A. & Vederas, J. Bacterial diaminopimelate metabolism as a target for antibiotic design. *Bioorg Med Chem* **8**, 843–871 (2000).
121. Hutton, C. A., Perugini, M. A. & Gerrard, J. A. Inhibition of lysine biosynthesis: an evolving antibiotic strategy. *Mol. Biosyst.* **3**, 458–465 (2007).
122. Paiva, A. *et al.* Inhibitors of dihydrodipicolinate reductase, a key enzyme of the diaminopimelate pathway of Mycobacterium tuberculosis. *Bba-Protein Struct M* **1545**, 67–77 (2001).
123. Lu, Y., Shevtchenko, T. N. & Paulus, H. Fine-structure mapping of cis-acting control sites in the lysC operon of Bacillus subtilis. *FEMS Microbiol. Lett.* **71**, 23–27 (1992).

124. Kochhar, S. & Paulus, H. Lysine-induced premature transcription termination in the lysC operon of *Bacillus subtilis*. *Microbiology (Reading, Engl)* **142** (Pt 7), 1635–1639 (1996).
125. Edwards, A. L., Garst, A. D. & Batey, R. T. Determining structures of RNA aptamers and riboswitches by X-ray crystallography. *Methods Mol Biol* **535**, 135–163 (2009).
126. Emsley, P. & Cowtan, K. Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* **60**, 2126–2132 (2004).
127. Ennifar, E., Walter, P., Ehresmann, B., Ehresmann, C. & Dumas, P. Crystal structures of coaxially stacked kissing complexes of the HIV-1 RNA dimerization initiation site. *Nat Struct Biol* **8**, 1064–1068 (2001).
128. Guo, F. & Cech, T. R. Evolution of *Tetrahymena* ribozyme mutants with increased structural stability. *Nat Struct Biol* **9**, 855–861 (2002).
129. Leontis, N. B., Lescoute, A. & Westhof, E. The building blocks and motifs of RNA architecture. *Curr Opin Struct Biol* **16**, 279–287 (2006).
130. Correll, C. C. *et al.* Crystal structure of the ribosomal RNA domain essential for binding elongation factors. *Proc Natl Acad Sci USA* **95**, 13436–13441 (1998).
131. Putnam, C. D., Hammel, M., Hura, G. L. & Tainer, J. A. X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* **40**, 191–285 (2007).
132. Lipfert, J. & Doniach, S. Small-angle X-ray scattering from RNA, proteins, and protein complexes. *Annual Reviews* (2007).
133. Russell, R. *et al.* Rapid compaction during RNA folding. *Proc Natl Acad Sci USA* **99**, 4266–4271 (2002).
134. Montange, R. K. *et al.* Discrimination between closely related cellular metabolites by the SAM-I riboswitch. *J Mol Biol* **396**, 761–772 (2010).
135. Lu, C. *et al.* Crystal structures of the SAM-III/S(MK) riboswitch reveal the SAM-dependent translation inhibition mechanism. *Nat Struct Mol Biol* **15**, 1076–1083 (2008).
136. Garst, A. D. & Batey, R. T. A switch in time: detailing the life of a riboswitch. *Biochim Biophys Acta* **1789**, 584–591 (2009).
137. Roth, A. & Breaker, R. R. The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem* **78**, 305–334 (2009).
138. Zhang, J., Lau, M. W. & Ferré-D'Amaré, A. R. Ribozymes and Riboswitches: Modulation of RNA Function by Small Molecules. *Biochemistry* **49**, 9123–9131 (2010).
139. Artsimovitch, I. & Landick, R. C. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci USA* **97**, 7090–7095 (2000).
140. Ha, K. S., Touloukhonov, I., Vassilyev, D. G. & Landick, R. C. The NusA N-terminal domain is necessary and sufficient for enhancement of transcriptional pausing via interaction with the RNA exit channel of RNA polymerase. *J Mol Biol* **401**, 708–725 (2010).

141. Yakhnin, A. V., Yakhnin, H. & Babitzke, P. Function of the *Bacillus subtilis* transcription elongation factor NusG in hairpin-dependent RNA polymerase pausing in the *trp* leader. *Proc Natl Acad Sci USA* **105**, 16131–16136 (2008).
142. Grundy, F. J. & Henkin, T. M. Kinetic analysis of tRNA-directed transcription antitermination of the *Bacillus subtilis* *glyQS* gene in vitro. *J Bacteriol* **186**, 5392–5399 (2004).
143. Garst, A. D., Heroux, A., Rambo, R. P. & Batey, R. T. Crystal structure of the lysine riboswitch regulatory mRNA element. *J Biol Chem* **283**, 22347–22351 (2008).
144. Grundy, F. J., Lehman, S. C. & Henkin, T. M. The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes. *Proc Natl Acad Sci USA* **100**, 12057–12062 (2003).
145. Baird, N. J. *et al.* Extended structures in RNA folding intermediates are due to nonnative interactions rather than electrostatic repulsion. *J Mol Biol* **397**, 1298–1306 (2010).
146. Watson, P. Y. & Fedor, M. J. The *glmS* riboswitch integrates signals from activating and inhibitory metabolites in vivo. *Nat Struct Mol Biol* **18**, 359–363 (2011).
147. Bennett, B. D. *et al.* Absolute metabolite concentrations and implied enzyme active site occupancy in *Escherichia coli*. *Nat Methods* **5**, 593–599 (2009).
148. Gherghe, C. M., Shajani, Z., Wilkinson, K. A., Varani, G. & Weeks, K. M. Strong correlation between SHAPE chemistry and the generalized NMR order parameter (S2) in RNA. *J Am Chem Soc* **130**, 12244–12245 (2008).
149. Theobald, D. L. & Wuttke, D. S. THESEUS: maximum likelihood superpositioning and analysis of macromolecular structures. *Bioinformatics* **22**, 2171–2172 (2006).
150. Gilbert, S. D., Reyes, F. E., Edwards, A. L. & Batey, R. T. Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. *Structure* **17**, 857–868 (2009).
151. Breaker, R. R. Prospects for riboswitch discovery and analysis. *Mol Cell* **43**, 867–879 (2011).
152. Goodrich, J. A. & Kugel, J. F. *Binding and Kinetics for Molecular Biologists*. 182 (Cold Spring Harbor Laboratory Press: 2007).
153. Korzheva, N. A Structural Model of Transcription Elongation. *Science* **289**, 619–625 (2000).
154. Peters, J. M., Vangeloff, A. D. & Landick, R. C. Bacterial Transcription Terminators: The RNA 3'-End Chronicles. *J Mol Biol* **412**, 793–813 (2011).
155. Proshkin, S., Rahmouni, A. R., Mironov, A. & Nudler, E. Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science* **328**, 504–508 (2010).
156. Vogel, U. & Jensen, K. F. The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *J Bacteriol* **176**, 2807–2813 (1994).

157. Zhou, J., Ha, K. S., La Porta, A., Landick, R. C. & Block, S. M. Applied force provides insight into transcriptional pausing and its modulation by transcription factor NusA. *Mol Cell* **44**, 635–646 (2011).
158. Taniguchi, Y. *et al.* Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. *Science* **329**, 533–538 (2010).
159. Artsimovitch, I. & Henkin, T. M. In vitro approaches to analysis of transcription termination. *Methods* **47**, 37–43 (2009).
160. Landick, R. C., Wang, D. & Chan, C. L. Quantitative analysis of transcriptional pausing by Escherichia coli RNA polymerase: his leader pause site as paradigm. *Meth Enzymol* **274**, 334–353 (1996).
161. Hyeon, C. & Thirumalai, D. Chain Length Determines the Folding Rates of RNA. (2012).
162. Poulsen, P. & Jensen, K. F. Effect of UTP and GTP pools on attenuation at the pyrE gene of Escherichia coli. *Mole Gen Genet* **208**, 152–158 (1987).
163. Barbolina, M. V., Kristoforov, R., Manfredo, A., Chen, Y. & Gollnick, P. The Rate of TRAP Binding to RNA Is Crucial for Transcription Attenuation Control of the B. subtilis trp Operon. *J Mol Biol* **370**, 925–938 (2007).
164. Wickiser, J. K. Kinetics of riboswitch regulation studied by in vitro transcription. *Methods Mol Biol* **540**, 53–63 (2009).
165. Blouin, S., Chinnappan, R. & Lafontaine, D. A. Folding of the lysine riboswitch: importance of peripheral elements for transcriptional regulation. *Nucleic Acids Res* (2010).doi:10.1093/nar/gkq1247
166. Smolke, C. D. Design Principles for Riboswitch Function. *PLoS Comput Biol* **5**, (2009).
167. Perdrizet, G. A., Artsimovitch, I., Furman, R., Sosnick, T. R. & Pan, T. Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proceedings of the National Academy of Sciences* **109**, 3323–3328 (2012).
168. Herbert, K. M. *et al.* Sequence-Resolved Detection of Pausing by Single RNA Polymerase Molecules. *Cell* **125**, 1083–1094 (2006).
169. McGraw, A. P., Bevilacqua, P. C. & Babitzke, P. TRAP-5' stem loop interaction increases the efficiency of transcription termination in the Bacillus subtilis trpEDCFBA operon leader region. *RNA* **13**, 2020–2033 (2007).
170. Merino, E. & Yanofsky, C. Transcription attenuation: a highly conserved regulatory strategy used by bacteria. *Trends Genet* **21**, 260–264 (2005).
171. Turnbough, C. L. & Switzer, R. L. Regulation of Pyrimidine Biosynthetic Gene Expression in Bacteria: Repression without Repressors. *Microbiology and Molecular Biology Reviews* **72**, 266–300 (2008).
172. Gaal, T., Bartlett, M. S., Ross, W., Turnbough, C. L. & Gourse, R. L. Transcription regulation by initiating NTP concentration: rRNA synthesis in bacteria. *Science* **278**, 2092–2097 (1997).
173. Garman, E. & Sweet, R. M. X-ray data collection from macromolecular crystals. *Methods Mol Biol* **364**, 63–94 (2007).

174. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* **66**, 213–221 (2010).
175. Benner, S. A. & Sismour, A. M. Synthetic biology. *Nat Rev Genet* **6**, 533–543 (2005).
176. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. *Nat Rev Genet* **11**, 367–379 (2010).
177. Isaacs, F. J. & Collins, J. J. Plug-and-play with RNA. *Nat Biotechnol* **23**, 306–307 (2005).
178. Liu, C. C. & Arkin, A. P. Cell biology. The case for RNA. *Science* **330**, 1185–1186 (2010).
179. Isaacs, F. J. *et al.* Engineered riboregulators enable post-transcriptional control of gene expression. *Nat Biotechnol* **22**, 841–847 (2004).
180. Callura, J. M., Dwyer, D. J., Isaacs, F. J., Cantor, C. R. & Collins, J. J. Tracking, tuning, and terminating microbial physiology using synthetic riboregulators. *Proc Natl Acad Sci USA* **107**, 15898–15903 (2010).
181. Lucks, J. B., Qi, L., Mutalik, V. K., Wang, D. & Arkin, A. P. Versatile RNA-sensing transcriptional regulators for engineering genetic networks. *Proceedings of the National Academy of Sciences* **108**, 8617–8622 (2011).
182. Qi, L., Lucks, J. B., Liu, C. C., Mutalik, V. K. & Arkin, A. P. Engineering naturally occurring trans-acting non-coding RNAs to sense molecular signals. *Nucleic Acids Res* (2012).doi:10.1093/nar/gks168
183. Friedland, A. E. *et al.* Synthetic gene networks that count. *Science* **324**, 1199–1202 (2009).
184. Tuerk, C. & Gold, L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* **249**, 505–510 (1990).
185. Ellington, A. D. & Szostak, J. W. In vitro selection of RNA molecules that bind specific ligands. *Nature* **346**, 818–822 (1990).
186. Vinkenborg, J. L., Karnowski, N. & Famulok, M. Aptamers for allosteric regulation. *Nat Chem Biol* **7**, 519–527 (2011).
187. Soukup, G. A. & Breaker, R. R. Engineering precision RNA molecular switches. *Proc Natl Acad Sci USA* **96**, 3584–3589 (1999).
188. Robertson, M. P. & Ellington, A. D. In vitro selection of an allosteric ribozyme that transduces analytes to amplicons. *Nat Biotechnol* **17**, 62–66 (1999).
189. Hesselberth, J. R., Robertson, M. P., Knudsen, S. M. & Ellington, A. D. Simultaneous detection of diverse analytes with an aptazyme ligase array. *Anal Biochem* **312**, 106–112 (2003).
190. Win, M. N. & Smolke, C. D. Higher-order cellular information processing with synthetic RNA devices. *Science* **322**, 456–460 (2008).
191. Sinha, J., Reyes, S. J. & Gallivan, J. P. Reprogramming bacteria to seek and destroy an herbicide. *Nat Chem Biol* **6**, 464–470 (2010).
192. Zhang, J., Lau, M. & Ferre-D'Amare, A. Ribozymes and Riboswitches: Modulation of RNA Function by Small Molecules. *Biochemistry*

193. Link, K. H. & Breaker, R. R. Engineering ligand-responsive gene-control elements: lessons learned from natural riboswitches. *Gene Ther.* **16**, 1189–1201 (2009).
194. Garst, A. D., Edwards, A. L. & Batey, R. T. Riboswitches: Structures and Mechanisms. *Cold Spring Harbor perspectives in biology* (2010).doi: 10.1101/cshperspect.a003533
195. Anthony, P. C., Perez, C. F., García-García, C. & Block, S. M. Folding energy landscape of the thiamine pyrophosphate riboswitch aptamer. *Proceedings of the National Academy of Sciences* **109**, 1485–1489 (2012).
196. Hofacker, I. L. Vienna RNA secondary structure server. *Nucleic Acids Res* **31**, 3429–3431 (2003).
197. Zuker, M. & Stiegler, P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res* **9**, 133–148 (1981).
198. Isambert, H. The jerky and knotty dynamics of RNA. *Methods* **49**, 189–196 (2009).
199. Hofacker, I., Maurer-Stroh, S., Stadler, P. & Zehl, M. Design of multistable RNA molecules. *RNA* (2001).
200. Gilbert, S. D., Love, C. E., Edwards, A. L. & Batey, R. T. Mutational analysis of the purine riboswitch aptamer domain. *Biochemistry* **46**, 13297–13309 (2007).
201. Lynch, S. A., Desai, S. K., Sajja, H. K. & Gallivan, J. P. A high-throughput screen for synthetic riboswitches reveals mechanistic insights into their function. *Chem Biol* **14**, 173–184 (2007).
202. Feng, X., Liu, L., Duan, X. & Wang, S. An engineered riboswitch as a potential gene-regulatory platform for reducing antibacterial drug resistance. *Chem Commun (Camb)* **47**, 173–175 (2011).
203. Ausländer, S., Ketzer, P. & Hartig, J. S. A ligand-dependent hammerhead ribozyme switch for controlling mammalian gene expression. *Mol. BioSyst.* **6**, 807–814 (2010).
204. Trausch, J. J., Ceres, P., Reyes, F. E. & Batey, R. T. The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer. *Structure* **19**, 1413–1423 (2011).
205. Win, M. N. & Smolke, C. D. A modular and extensible RNA-based gene-regulatory platform for engineering cellular function. *Proc Natl Acad Sci USA* **104**, 14283–14288 (2007).
206. Crick, F. H. The origin of the genetic code. *J Mol Biol* **38**, 367–379 (1968).
207. Benner, S. A., Ellington, A. D. & Tauer, A. Modern metabolism as a palimpsest of the RNA world. *Proc Natl Acad Sci USA* **86**, 7054–7058 (1989).
208. Jeffares, D. C., Poole, A. M. & Penny, D. Relics from the RNA world. *J Mol Evol* **46**, 18–36 (1998).
209. Yarus, M. Getting past the RNA world: the initial Darwinian ancestor. *Cold Spring Harbor perspectives in biology* **3**, (2011).

210. Knight, R. Finding specific RNA motifs: Function in a zeptomole world? *RNA* **9**, 218–230 (2003).
211. Turk, R. M., Chumachenko, N. V. & Yarus, M. Multiple translational products from a five-nucleotide ribozyme-SUPPLEMENT. *Proc Natl Acad Sci USA* **107**, 4585–4589 (2010).
212. Cech, T. R. Crawling out of the RNA world. *Cell* **136**, 599–602 (2009).
213. Blouin, S., Mulhbachter, J., Penedo, J. & Lafontaine, D. Riboswitches: Ancient and Promising Genetic Regulators. *ChemBioChem* (2008).doi: 10.1002/cbic.200800593
214. Ribas de Pouplana, L. & Schimmel, P. Two classes of tRNA synthetases suggested by sterically compatible dockings on tRNA acceptor stem. *Cell* **104**, 191–193 (2001).
215. Onesti, S. *et al.* Structural studies of lysyl-tRNA synthetase: conformational changes induced by substrate binding. *Biochemistry* **39**, 12853–12861 (2000).
216. Levensgood, J., Ataide, S. F., Roy, H. & Ibba, M. Divergence in noncognate amino acid recognition between class I and class II lysyl-tRNA synthetases. *J Biol Chem* **279**, 17707–17714 (2004).
217. Ataide, S. F. *et al.* Mechanisms of resistance to an amino acid antibiotic that targets translation. *ACS Chem Biol* **2**, 819–827 (2007).
218. Cochrane, J. C. & Strobel, S. A. Riboswitch effectors as protein enzyme cofactors. *RNA* **14**, 993–1002 (2008).
219. Fedor, M. J. & Williamson, J. R. The catalytic diversity of RNAs. *Nat Rev Mol Cell Biol* **6**, 399–412 (2005).
220. Patte, J. & Akrim, M. The leader sequence of the Escherichia coli lysC gene is involved in the regulation of LysC synthesis - Patte - 2006 - FEMS Microbiology Letters - Wiley Online Library. *FEMS Microbiol. Lett.* (1998).
221. Sudarsan, N., Cohen-Chalamish, S., Nakamura, S., Emilsson, G. M. & Breaker, R. R. Thiamine pyrophosphate riboswitches are targets for the antimicrobial compound pyrithiamine. *Chem Biol* **12**, 1325–1335 (2005).
222. Lee, E. R., Blount, K. F. & Breaker, R. R. Roseoflavin is a natural antibacterial compound that binds to FMN riboswitches and regulates gene expression. *RNA Biol* **6**, 187–194 (2009).