CHARACTERIZATION OF THE LYSINE-RESPONSIVE L BOX RIBOSWITCH

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy in the Graduate School of The Ohio State University

By

Sharnise Nicole Mitchell

The Ohio State Biochemistry Graduate Program

The Ohio State University

2012

Dissertation Committee:

Professor Tina Henkin, Advisor

Professor Juan Alfonzo

Professor Don Dean

Professor Mark Foster
ABSTRACT

Regulation of gene expression is an essential process that organisms employ for growth and survival. The activation or repression of genes is critical as it allows the cell to monitor environmental signals and adapt to changes in nutrient availability or environmental conditions. Regulation can occur at any step during gene expression and there are a number of mechanisms employed. One method of regulation involves RNA elements termed riboswitches. Riboswitches are conserved RNA elements commonly located upstream of the gene under regulatory control. These RNAs regulate gene expression by modulation of the RNA structure. This method of gene regulation requires only the RNA and regulatory signal without the need for any additional cellular factors. Typically, a structural change in the RNA occurs in response to environmental signals such as temperature, small RNAs, or small molecules. Most of the known riboswitches modulate gene expression in response to a small molecule effector. Recognition of the effector causes a structural rearrangement that prevents or promotes the formation of a regulatory structure such as an intrinsic transcriptional terminator. Riboswitches that regulate gene expression at the level of translation undergo a structural rearrangement that can occlude or expose the ribosomal binding site (RBS).
In this work, we focused our analysis on the L box riboswitch that regulates gene expression during transcription in *Bacillus subtilis* and is predicted to regulate at the translational level in *Escherichia coli*. This conserved RNA regulates the expression of lysine biosynthetic genes in response to cellular lysine concentrations. Previous work revealed that the *lysC* leader RNA from *B. subtilis* promotes premature transcription termination in the presence of lysine and can discriminate against lysine analogs. Analysis of the leader region *in vivo* indicated that *lysC* expression is repressed when the RNA is transcribed in the presence of high lysine. The current model of the *lysC* structural rearrangement that modulates gene expression predicts that the leader RNA binds lysine, which results in stabilization of the ligand-binding and terminator structures. In the absence of lysine, the leader RNA forms an alternate antiterminator that prevents formation of the terminator.

Here we investigate the conserved sequence and structural features of the *lysC* leader riboswitch to elucidate the features that are required for lysine binding and the required structural transition. We have identified variants of the *lysC* leader RNA that stabilize the termination conformation in the absence of lysine. These results suggest that the energetic balance between the termination and antitermination conformations is sequence-dependent and the conserved nucleotides are critical to the stabilization of the ligand-free structure. *In vitro* evolution of a randomly mutated *lysC* leader indicated that regions with conserved structural motifs are also critical to the lysine sensitivity and the structural transition required for lysine-dependent repression. Mutation of non-conserved
residues demonstrated that regions distal to the ligand-binding pocket affect ligand sensitivity. Finally, site-directed mutagenesis was used to identify the features of the lysine-binding pocket that are responsible for the lysine specificity. Our results indicate that the majority of the conserved nucleotides within the binding pocket are required for recognition of independent features of the lysine molecule. Mutation of the conserved residues yielded lysC leader variants with alternate ligand specificities. These mutations resulted in variants that are specific for lysine analogs and exhibit a loss of lysine recognition. These experiments provide a comprehensive view of the lysC leader RNA requirements for gene regulation and highlight the conserved sequence and structural features that are necessary for function of the L box riboswitch.
This document is dedicated to my family,
especially my husband, Marlon,
and my parents, Stanford and Sylvia Wilson.
ACKNOWLEDGEMENTS

I would like to express my sincere gratitude and appreciation to my advisor Dr. Tina Henkin. I am extremely grateful for all of the time, effort, and guidance that she has given to support my scientific aspirations and career goals.

I wish to thank Dr. Frank Grundy for his support and encouragement throughout this process. Dr. Grundy played an instrumental role in this project and was particularly important to the \textit{in vitro} evolution of the \textit{lysC} leader RNA. I would also like to thank Dr. Mike Ibba and the Ibba lab for including the \text{L box riboswitch} in their work on LysRS2, and for providing the lysine analogs.

To my committee members Dr. Juan Alfonzo, Dr. Don Dean, and Dr. Mark Foster I am truly grateful to have had the opportunity to work with such an amazing group of scientists. I really appreciate your support and guidance throughout the years.

I am extremely grateful to my wonderful family and friends that have been an endless source of support, encouragement, and strength. In particular, I would like to thank my parents, for their countless words of encouragement and undying support. I would like to thank my husband, my dearest friend, and life partner; I could not have done this without you. I wish to thank Terri, Sydika, Ursula, Betina, Vineeta, Victoria, and Kim. Thank each of you for taking this journey with me. But most importantly, I must give all of my thanks to God, without whom none of this work would be possible.
VITA

April 13, 1981 ........................................Born-Los Angeles, CA

2003..............................................................B.S. Chemistry, Tennessee State University

2003 to present ...........................................Graduate Research Associate, Ohio State

Biochemistry Program, The Ohio State University

PUBLICATIONS


FIELDS OF STUDY

Major Field: Biochemistry
# TABLE OF CONTENTS

**ABSTRACT** .................................................................................................................................................. ii

**ACKNOWLEDGEMENTS** .............................................................................................................................. vi

**VITA** ............................................................................................................................................................ vii

**PUBLICATIONS** ........................................................................................................................................ vii

**FIELDS OF STUDY** .................................................................................................................................... vii

**LIST OF TABLES** ...................................................................................................................................... xi

**LIST OF FIGURES** ...................................................................................................................................... xii

**LIST OF ABBREVIATIONS** ........................................................................................................................ xiv

**CHAPTER 1** ............................................................................................................................................... 1

**GENETIC REGULATION BY RIBOSWITCHES** ......................................................................................... 1

1.1 Riboswitch classes .................................................................................................................................. 5

1.1.1 Temperature and pH-sensing riboswitches ....................................................................................... 5

1.1.2 The T box riboswitch ....................................................................................................................... 8

1.1.3 Purine-sensing riboswitches ........................................................................................................... 12

1.1.4 Cyclic-di-GMP riboswitches .......................................................................................................... 19

1.1.5 The M box riboswitch ................................................................................................................... 22

1.1.6 The *glmS* ribozyme .................................................................................................................... 24

1.1.7 Coenzyme-sensing riboswitches ..................................................................................................... 26

1.1.8 S box riboswitch ............................................................................................................................ 37

1.1.9 Amino acid-sensing riboswitches .................................................................................................... 50

1.2 Goals of this study .................................................................................................................................. 61

**CHAPTER 2** .............................................................................................................................................. 64
Mutational analysis of L box structural elements ................................................................. 64

2.1 Introduction .......................................................................................................................... 64

2.2 Materials and Methods ....................................................................................................... 70
  2.2.1 Generation of DNA templates ...................................................................................... 70
  2.2.2 Site-directed mutagenesis ......................................................................................... 74
  2.2.3 RNase H cleavage assay ........................................................................................... 74
  2.2.4 *In vitro* transcription ................................................................................................. 75
  2.2.5 *lysC-lacZ* transcriptional fusions and β-galactosidase assays .............................. 76

2.3 Results ................................................................................................................................. 77
  2.3.1 Lysine promotes stabilization of helix 1 .................................................................... 77
  2.3.2 Mutational analysis of helix 1 .................................................................................... 81
  2.3.3 *In vivo* expression of variant *lysC-lacZ* transcriptional fusions ...................... 87
  2.3.4 Mutational analysis of helix 4 terminal loop ........................................................... 90

2.4 Discussion ............................................................................................................................ 93

CHAPTER 3 .................................................................................................................................. 101

*In vitro* evolution of variant L box leader RNAs ................................................................. 101

3.1 Introduction .......................................................................................................................... 101

3.2 Materials and Methods ....................................................................................................... 105
  3.2.1 Construction of DNA templates .............................................................................. 105
  3.2.2 RNA synthesis and purification .............................................................................. 108
  3.2.3 Lysine-binding assay ............................................................................................... 108
  3.2.4 Fluorescence spectroscopy ...................................................................................... 109
  3.2.5 *In vitro* selection using SELEX ............................................................................ 110
  3.2.6 RNase H cleavage assay ......................................................................................... 113
  3.2.7 *In vitro* transcription of SELEX isolated variants .............................................. 114
  3.2.8 *lysC-lacZ* transcriptional fusions and β-galactosidase assays ........................... 114

3.3 Results .................................................................................................................................. 115
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>3.3.1</td>
<td>Analysis of lysine binding by the <em>B. subtilis</em> <em>lysC</em> leader RNA</td>
<td>115</td>
</tr>
<tr>
<td>3.3.2</td>
<td>Design of a circularly permutated bipartite RNA with a 2-AP substitution</td>
<td>117</td>
</tr>
<tr>
<td>3.3.3</td>
<td>Analysis of lysine binding using spectrofluorimetry</td>
<td>121</td>
</tr>
<tr>
<td>3.3.4</td>
<td>Design of SELEX selection scheme</td>
<td>124</td>
</tr>
<tr>
<td>3.3.5</td>
<td>Identification and characterization of SELEX variants</td>
<td>127</td>
</tr>
<tr>
<td>3.3.6</td>
<td><em>In vitro</em> selection of variant pool using a modified SELEX scheme</td>
<td>136</td>
</tr>
<tr>
<td>3.4</td>
<td>Discussion</td>
<td>141</td>
</tr>
<tr>
<td>CHAPTER 4</td>
<td></td>
<td>147</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>147</td>
</tr>
<tr>
<td>4.2</td>
<td>Materials and Methods</td>
<td>154</td>
</tr>
<tr>
<td>4.2.1</td>
<td>Construction of DNA templates</td>
<td>154</td>
</tr>
<tr>
<td>4.2.2</td>
<td>Site-directed Mutagenesis</td>
<td>156</td>
</tr>
<tr>
<td>4.2.3</td>
<td><em>In vitro</em> transcription termination assays</td>
<td>157</td>
</tr>
<tr>
<td>4.2.4</td>
<td>RNase H cleavage</td>
<td>158</td>
</tr>
<tr>
<td>4.3</td>
<td>Results</td>
<td>159</td>
</tr>
<tr>
<td>4.3.1</td>
<td>Wild-type <em>lysC</em> transcription termination in response to lysine analogs</td>
<td>159</td>
</tr>
<tr>
<td>4.3.2</td>
<td>Effects of substitutions at G111 and A112</td>
<td>162</td>
</tr>
<tr>
<td>4.3.3</td>
<td>Mutations near the lys carboxyl group result in altered ligand specificity</td>
<td>165</td>
</tr>
<tr>
<td>4.3.4</td>
<td>$K^+$ decreases the concentration of lys required for transcription termination.</td>
<td>170</td>
</tr>
<tr>
<td>4.3.5</td>
<td><em>lysC</em> leader RNA variants undergo ligand-dependent structural changes</td>
<td>172</td>
</tr>
<tr>
<td>4.4</td>
<td>Discussion</td>
<td>175</td>
</tr>
<tr>
<td>CHAPTER 5</td>
<td></td>
<td>180</td>
</tr>
<tr>
<td>5.1</td>
<td>Summary and Future directions</td>
<td>180</td>
</tr>
<tr>
<td>REFERENCES</td>
<td></td>
<td>189</td>
</tr>
</tbody>
</table>
LIST OF TABLES

Table

2.1. DNA oligonucleotide primers for the *B. subtilis* lysC leader .................................. 72

2.2. DNA oligonucleotide primers for the *E. coli* lysC leader ..................................... 73

2.3. Expression of *lysC-lacZ* fusions in *B. subtilis* .................................................... 88

3.1. DNA oligonucleotides for circularly permuted lysC leader RNA synthesis ........ 107

3.2. lysC leader RNA variants identified by *in vitro* selection. ................................. 129

3.3. lysC leader variants isolated from additional *in vitro* selection ...................... 141

4.1. *In vitro* transcription termination of wild-type lysC leader in the presence of lys or lys analogs............................................................................................................. 162

4.2. *In vitro* transcription termination of G111 and A112 lysC leader variants in the presence of lys or lys analogs .......................................................... 164

4.3. *In vitro* transcription termination of G40-C110 lysC leader variants in the presence of 10 mM lys or lys analogs. ............................................................. 167

4.4. Termination efficiency of lysC leader variants in the presence of lys or lys analogs... ...................................................................................................................... 168

4.5. *In vitro* transcription termination of G144-U170 lysC leader variants in the presence of 10 mM lys or lys analogs. ............................................................. 170
LIST OF FIGURES

Figure

1.1. Example of generic thermo-sensing riboswitch.............................................................. 6
1.2. The T box riboswitch..................................................................................................... 11
1.3. Secondary structures of representative purine-sensing riboswitches......................... 15
1.4. Structure of the c-di-GMP binding pocket from \textit{V. cholera}................................. 22
1.5. X-ray crystal structure of the TPP bound aptamer domain........................................ 29
1.6. Structural models of the RFN element........................................................................... 33
1.7. Structures of the \textit{F. nucleatum impX} RFN element bound to FMN and analogs. .... 34
1.8. S box riboswitch. .......................................................................................................... 39
1.9. Structure of S\textsubscript{MK} box riboswitch. ............................................................... 44
1.10. Crystal structure of SAH binding pocket......................................................................... 49
1.11. Glycine riboswitch aptamers.......................................................................................... 51
1.12. Sequence and predicted secondary structural model of the \textit{B. subtilis lysC} leader RNA.................................................................................................................. 57
1.13. Global RNA structure of \textit{T. maritima} L box aptamer domain................................. 59
1.14. Interactions of lysine within the lysine-binding pocket .............................................. 60
2.1. Lysine-dependent structural transition............................................................................ 66
2.2. Helix 1 cleavage of the \textit{B. subtilis lysC} leader RNA by RNase H............................... 78
2.3. Helix 1 cleavage of the E. coli lysC leader RNA by RNase H. ................................. 80
2.4. In vitro transcription of U35-A198 variant lysC leader RNAs. ................................. 83
2.5. In vitro transcription of A36-U197 variant lysC leader RNAs. ................................. 86
2.6. In vitro transcription of A156 and A157 lysC leader RNA variants. ........................ 92

3.1. Lysine binding to the B. subtilis and E. coli lysC leader RNAs. ............................. 117
3.2. Bipartite RNA used to monitor lysine binding in fluorescence assays. .................. 119
3.3. Fluorescent emission of 2-AP labeled bipartite leader RNA. ................................. 123
3.4. In vitro selection scheme. ....................................................................................... 126
3.5. RNase H cleavage of lysC leader variants identified by SELEX. .............................. 131
3.6. In vitro transcription of Class II and Class III SELEX variants. .............................. 133
3.7. Lysine-dependent expression of wild type and mutant lysC-lacZ fusions. .............. 136
3.8. Modified SELEX scheme. ....................................................................................... 140

4.1. Chemical structures of lysine and closely related lysine analogs. ......................... 150
4.2. Tertiary structure of the Thermatoga maritima L box lysine-binding pocket. ....... 153
4.3. In vitro transcription of wild type and variant lysC leader RNAs in the presence and absence of potassium. ................................................................. 171
4.4. RNase H cleavage of wild type and variant lysC leader RNAs. ............................. 173
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>2-AP</td>
<td>2-aminopurine</td>
</tr>
<tr>
<td>2’-OH</td>
<td>2’-hydroxyl</td>
</tr>
<tr>
<td>Å</td>
<td>angstrom</td>
</tr>
<tr>
<td>aa</td>
<td>amino acid</td>
</tr>
<tr>
<td>AASD</td>
<td>anti-anti-Shine-Dalgarno</td>
</tr>
<tr>
<td>aaRS</td>
<td>aminoacyl-tRNA synthetase</td>
</tr>
<tr>
<td>AAT</td>
<td>anti-antiterminator</td>
</tr>
<tr>
<td>AdoCbl</td>
<td>5’-deoxy 5’-adenosylcobalamin</td>
</tr>
<tr>
<td>AEC</td>
<td>aminoethylcysteine</td>
</tr>
<tr>
<td>ASD</td>
<td>anti-Shine-Dalgarno</td>
</tr>
<tr>
<td>AT</td>
<td>antiterminator</td>
</tr>
<tr>
<td>ATP</td>
<td>adenosine triphosphate</td>
</tr>
<tr>
<td>bp</td>
<td>base pair</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>c-di-GMP</td>
<td>cyclic diguanylate</td>
</tr>
<tr>
<td>cDNA</td>
<td>complementary DNA</td>
</tr>
<tr>
<td>CTP</td>
<td>cytidine triphosphate</td>
</tr>
<tr>
<td>DAP</td>
<td>diaminopimelate</td>
</tr>
</tbody>
</table>
dG  deoxyguanine
D-lys  D-lysine
DNA  deoxyribonucleic acid
dNTP  deoxynucleotide triphosphate
FMN  flavin mononucleotide
GEMM  genes for the environment, membranes and motility
GlcN6P  glucosamine-6-phosphate
GTP  guanosine triphosphate
hydroxy lys  DL-5-hydroxylysine
$K_D$  dissociation constant
kDa  kilodalton
L-arg  L-arginine
lysEE  lysine ethyl ester
lysOH  lysine hydroxamate
lysNH  lysinamide
lys  lysine
lys methyl  lysine methyl
LysRS  lysyl-tRNA synthetase
mRNA  messenger RNA
NAD$^+$  nicotinamide adenine dinucleotide
NMIA  N-methylisatoic anhydride
nt  nucleotide
<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>NTP</td>
<td>nucleotide triphosphate</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>preQ&lt;sub&gt;1&lt;/sub&gt;</td>
<td>7-aminoethly-7-deazaguanine</td>
</tr>
<tr>
<td>RBS</td>
<td>ribosome binding-site</td>
</tr>
<tr>
<td>ROSE</td>
<td>repression of heat-shock gene expression</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>RNAP</td>
<td>RNA polymerase</td>
</tr>
<tr>
<td>RNase</td>
<td>ribonuclease</td>
</tr>
<tr>
<td>SAH</td>
<td>S-adenosylhomocysteine</td>
</tr>
<tr>
<td>SAM</td>
<td>S-adenosylmethionine</td>
</tr>
<tr>
<td>SD</td>
<td>Shine-Dalgarno</td>
</tr>
<tr>
<td>SHAPE</td>
<td>2’-hydroxyl acylation analyzed by primer extension</td>
</tr>
<tr>
<td>T</td>
<td>terminator</td>
</tr>
<tr>
<td>TPP</td>
<td>thiamin pyrophosphate</td>
</tr>
<tr>
<td>Tris-HCl</td>
<td>tris-(hydroxymethyl) aminoethane hydrochloride</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>Trm&lt;sub&gt;1/2&lt;/sub&gt;</td>
<td>half-maximal termination</td>
</tr>
<tr>
<td>UTP</td>
<td>uridine triphosphate</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
</tbody>
</table>
CHAPTER 1

GENETIC REGULATION BY RIBOSWITCHES

The central dogma of molecular biology describes gene expression as a linear process in which the genetic information stored in DNA is transcribed into mRNA. The mRNA is translated into proteins that perform the catalytic processes. Recent work has revealed that RNA performs essential processes once thought to be the responsibility of proteins. One such process is the regulation of gene expression, which allows organisms to adjust gene expression in response to environmental and metabolic changes. In bacteria, adaptability is extremely important as a cell often encounters environments that require survival in harsh conditions with limited nutrients. Thus, bacteria have adopted a number of sophisticated mechanisms to overcome these challenges to ensure that gene expression occurs under the appropriate conditions.

Recent characterization of some bacterial non-coding RNAs has revealed a number of genetic regulatory mechanisms (Smith et al., 2010a). One example involves the control of gene expression by riboswitches. Riboswitches are regions of conserved sequence and structural features embedded in the 5- untranslated region (UTR) of an mRNA, upstream of the coding sequence; this region is also termed the leader RNA.
Unlike other small RNAs that control expression in trans, riboswitch RNAs are primarily cis-acting control elements that regulate the expression of an adjacent coding sequence (Henkin, 2008). During riboswitch-mediated regulation, gene expression is modulated in response to a metabolic signal; this change in gene expression occurs without additional proteins or enzymatic factors. The recognized metabolic signal can be temperature, small RNAs, or small molecules (Henkin, 2008). Currently, the majority of riboswitches consist of two main functional regions, the aptamer domain which is responsible for recognition of the signal and the expression platform which directs gene expression (Barrick and Breaker, 2007). Signal recognition by the aptamer causes structural changes in the RNA that dictate the structure of the expression platform. The expression platform most commonly regulates gene expression by affecting the formation of an intrinsic (rho-independent) transcriptional terminator, or an mRNA structure that prevents translation initiation (Henkin, 2008).

Regulation of gene expression by riboswitches is carried out through structural changes in the RNA that govern the formation of two mutually exclusive structures. Premature transcription termination is one mechanism used to regulate gene expression through the competitive formation of terminator (T) or antiterminator (AT) elements (Henkin, 2008). Formation of the terminator stops the progression of transcription, whereas formation of the antiterminator helix allows transcription to continue. If the cellular requirement is transcription termination, sequences involved in antiterminator
formation are sequestered by an alternate anti-antiterminator (AAT) element; this allows formation of the terminator that prevents expression of the downstream gene. The regulated gene is often directly involved in the biosynthesis of the regulatory signal. If the concentration of the environmental signal or effector molecule is sufficient for growth and survival, then expression of the gene is unnecessary and may consume valuable resources. Alternatively, riboswitches can use the formation of mutually exclusive structures to induce expression of the downstream gene, which will allow the production of the effector molecule as needed.

Riboswitch RNAs can also regulate gene expression at the level of translation initiation. Translation of mRNAs requires binding of the translation initiation complex to the ribosome-binding site (RBS), which consists of the Shine-Dalgarno (SD) recognition sequence and AUG start codon of the open reading frame (ORF). Sequestration of the ribosomal recognition elements can occur through interactions with alternate sequences known as anti-Shine-Dalgarno (ASD) regions, which results in repression of gene expression. Under conditions when gene expression is advantageous, the ASD sequence can be paired with a complementary anti-anti-Shine-Dalgarno (AASD) sequence, and thus allow access of the ribosome to the SD. In contrast, some riboswitches contain a single RNA structure responsible for both signal recognition and regulation of gene expression (Lu et al., 2008).
Identification of riboswitches in eukaryotic organisms has uncovered regulatory control during mRNA splicing (Dambach and Winkler, 2009). Riboswitches of this type reside near an mRNA splice site, as opposed to upstream of the coding region. Binding of the cognate signal molecule promotes structural rearrangements that alter the availability of the adjacent splice site required for mRNA processing (Cheah et al., 2007). Occlusion of the splice site yields an alternate gene product that is targeted for degradation.

Current knowledge of riboswitch mechanisms illustrates that modulation of gene expression can occur during transcription, translation, and mRNA processing. Other riboswitch features such as tandem architectures and dual transcription/translational mechanisms have evolved to enhance function (Breaker, 2012). Although similar mechanisms are utilized to modulate gene expression, the effector molecules are extremely diverse. Characterization of a large number of bacterial riboswitches has revealed that these RNAs are widely distributed and control gene expression in response to a variety of signals. The thiamin pyrophosphate (TPP) riboswitch class contains at least one representative in all three domains of life (Bastet et al., 2011; Breaker, 2012). The current assortment of small molecule riboswitches and their pervasive function suggests that more riboswitch classes exist and await discovery, particularly in higher organisms.
1.1 Riboswitch classes

1.1.1 Temperature and pH-sensing riboswitches

The majority of known riboswitches regulate gene expression in response to a small RNA, small molecule, or metabolite. However, other RNAs exist that sense critical cellular parameters such as temperature and pH (Bastet et al., 2011). Control of gene expression in response to temperature occurs by a mechanism similar to that of metabolite sensing riboswitches. Most thermo-sensing riboswitches reside upstream of the gene under control and regulate gene expression during translation initiation (Narberhaus et al., 2006). Often, thermo-sensing RNAs direct the synthesis of proteins involved in the heat-shock response. At 30°C, translation of heat-shock proteins is prevented through the formation of an RNA hairpin that sequesters the RBS. A rapid increase in temperature to 42°C melts the hairpin and permits ribosome access to the RBS (Fig. 1.1) (Narberhaus et al., 2006; Rinnenthal et al., 2010). Additional studies of temperature sensing by these RNAs suggested that melting of the RNA helix is not the only mechanism utilized to monitor changes in temperature (Shah and Gilchrist, 2010). Analysis of helical melting rates demonstrated that RNAs not involved in thermo-sensing exhibited melting properties similar to the thermo-sensing RNAs. This suggests that there are other factors sensed by the RNA that support the temperature-dependent structural change. Additional studies are required to determine the specific RNA features that are responsible for the temperature sensitivity.
Figure 1.1. **Example of generic thermo-sensing riboswitch.** At low temperatures, formation of the hairpin structure prevents ribosome (light blue circles) binding to the RBS. When the temperature is increased, the Shine-Dalgarno (SD) and AUG start codon (dark blue square) are accessible for translation initiation. Large black arrow indicates progression of translation. Modified from Johansson *et al.* (2002).

While the mechanism of temperature sensing by the RNA thermo-sensors remains under debate, it is clear that RNAs have the unique ability to sense changes in temperature. For example, the *Escherichia coli rpoH* gene encodes a sigma factor (σ^{32}). This subunit of RNA polymerase (RNAP) is responsible for the recognition of promoters required for the transcription of genes that encode heat-shock proteins (Morita *et al.*, 1999b). Analysis of the *rpoH* mRNA identified temperature-dependent regulation that occurs at the translational level. At 30°C, *rpoH* expression is low; however, an increase in temperature to 42°C enhances translation. Structural probing indicated that the increase in temperature causes changes in the RNA structure that promote translation initiation and thus increase the cellular concentration of σ^{32}. This is consistent with another widely distributed class of thermo-sensing RNAs known as **Repression Of heat-**
Shock gene Expression (ROSE) elements (Narberhaus et al., 1998). ROSE elements can be found in the 5’ UTR of heat-shock genes in α- and γ-proteobacteria (Waldminghaus et al., 2005). ROSE elements repress gene expression under normal growth conditions and induce synthesis of heat-shock proteins in response to a rapid increase in temperature (Narberhaus et al., 1998). The E. coli cold-shock gene cspA (Giuliodori et al., 2010; Jiang et al., 1997) exhibits a response opposite to that of the heat-shock genes. CspA is an RNA chaperone utilized to prevent secondary structure stabilization at low temperatures (Jiang et al., 1997). Structural probing of the 5’ leader region of cspA revealed two temperature-dependent conformations. At 10°C, the cspA RNA adopts a structure that allows access of the ribosome to the RBS. However, at 37°C the RNA folds into an alternate structure that provides limited access for ribosomes to the RBS (Giuliodori et al., 2010). It was suggested that the decrease in temperature is responsible for the structural change that results in increased cspA translation (Jiang et al., 1997).

The prfA gene in Listeria monocytogenes is also controlled by changes in temperature (Leimeister-Wachter et al., 1990). prfA encodes a regulatory protein of the hemolytic toxin listeriolysin, a critical virulence factor of L. monocytogenes. At 30°C, PrfA is present in low concentrations due to RBS occlusion in the 5’ UTR (Johansson et al., 2002). A temperature increase to 37°C weakens interactions of the prfA mRNA and makes the RBS more accessible for translation (Johansson et al., 2002). Likewise, genes critical to the lysogenic cycle of bacterial phage λ, such as cIII, are also regulated by
RNA thermo-sensors. In this case, the cIII RNA can adopt two mutually exclusive temperature-dependent structures. At 37°C, the cIII RNA forms a structure that allows translation. However, a decrease in temperature causes a structural change that occludes access to the RBS (Altuvia et al., 1989).

Riboswitches can monitor chemical parameters such as pH, in addition to temperature changes. Translational fusions demonstrated that high pH conditions resulted in the up-regulation of the alx gene in E. coli (Nechooshtan et al., 2009). Deletion of the leader region caused constitutive expression of the translational fusion under alkaline conditions. Mutation of conserved residues predicted to interact when pH levels are high resulted in a loss of regulation. This work suggests that non-coding RNAs have evolved a chemical sensitivity that can monitor changes in cellular pH. Determination of the how the leader RNAs recognizes these chemical changes will be critical to understanding pH-sensing RNAs.

1.1.2 The T box riboswitch

The discovery and characterization of the Bacillus subtilis T box system began with a report that expression of the B. subtilis tyrS gene, which encodes tyrosyl-tRNA synthetase (TyrRS), increased in response to tyrosine limitation (Dale and Nester, 1971). Inspection of the tyrS leader region revealed a conserved 14-nucleotide sequence, designated the “T box,” upstream of an intrinsic terminator (Henkin et al., 1992). Mutation of the T box resulted in a reduction of tyrS expression and a loss of induction.
when tyrosine was limited. Phylogenetic analysis of sequence upstream of 9 other aminoacyl-tRNA synthetase (aaRS) genes uncovered a sequence and secondary structural pattern similar to that of the tyrS leader region. It was suggested that the leader RNA pattern was involved in a regulatory mechanism adapted to monitor the availability of amino acids as indicated by tRNA charging (Grundy and Henkin, 1993). Further analysis of the leader RNA sequences revealed a conserved codon designated the “Specifier Sequence” at a conserved position in the RNA structure. The codon corresponded to the amino acid substrate of the aaRS encoded by the downstream ORF. Substitution of the tyrosine codon in the tyrS leader with a phenylalanine codon resulted in induction of tyrS in response to phenylalanine limitation, supporting the model that the Specifier Sequence is responsible for the amino acid specific response. To date, >2,000 genes have been identified that contain T box riboswitches. The T box is highly represented in Gram-positive organisms with the most extensively studied T box riboswitches found upstream of the tyrS and glyQS genes in B. subtilis.

Genetic analysis of the tyrS leader RNA suggested that the tRNA was the effector molecule (Grundy and Henkin, 1993). The secondary structural model (Fig. 1.2) predicts that an uncharged tRNA can make two interactions with the leader RNA; one interaction ensures that the appropriate tRNA is bound through pairing between the Specifier Sequence and the anticodon, and the other stabilizes an antiterminator by pairing with the NCCA end of the tRNA. Stabilization of the antiterminator prevents formation of the terminator, as the same residues are required for formation of either structure (Fig. 1.2).
This mechanism provides a system for cells to monitor the charging ratios of specific tRNAs (Green et al., 2010). Binding of an aminoacylated tRNA by the leader RNA promotes formation of the terminator structure because the presence of the amino acid prevents the tRNA from interacting with the antiterminator helix. In contrast, when a non-aminoacylated tRNA is bound, interactions between the NCCA end and the antiterminator helix prevent formation of the terminator and allow transcription of the downstream ORF.
Figure 1.2. The T box riboswitch. Two critical interactions between the tRNA and the T box leader RNA determine the fate of the downstream gene. Under high tRNA charging conditions, the aminoacylated tRNA can interact with the leader RNA only at the Specifier Loop, which is responsible for tRNA recognition. The lack of a second interaction allows the formation of the terminator helix (blue and black helix) and prevents transcription of the downstream gene. Uncharged tRNA can bind at both the Specifier Sequence and the antiterminator, and the leader RNA undergoes structural modulation. The interaction of the uncharged tRNA stabilizes the antiterminator helix and allows transcription to continue. Green lines indicate base pairing interactions between leader RNA and tRNA. Blue and red lines indicate sequences involved in the formation of mutually exclusive terminator and antiterminator structures. The tRNA is shown in cyan and the yellow circle (aa) represents the amino acid. Adapted from Green et al. (2010).

Biochemical studies confirmed that uncharged tRNA promotes termination of the T box riboswitch (Grundy et al., 2002). Investigation of the tRNA requirements for antitermination supported the model in which two interactions between the tRNA and the leader RNA are essential. In vitro transcription of the B. subtilis glyQS leader RNA
the presence of variant tRNAs indicated that the full structure of the tRNA is essential for antitermination (Yousef et al., 2003). Additional structural analysis confirmed previously proposed interactions. The full-length leader RNA interacts directly with the tRNA in two places, the anticodon loop and CCA end (Yousef et al., 2005). Disruption of the D-loop/T-loop tertiary interaction in the tRNA resulted in a loss of antitermination. The addition of a base pair in the anticodon stem was tolerated however; further extension of the anticodon helix abolished antitermination. In vitro analysis also revealed that tRNA-dependent antitermination exhibits a face-of-the-helix dependence in which the NCCA end of the tRNA must be precisely positioned to interact with the antiterminator bulge. The discovery of the first RNA-sensing riboswitch describes a simple strategy that organisms utilize to monitor the charging ratios of cellular tRNAs. If the majority of a particular tRNA class is charged, aaRS production is unnecessary. The T box riboswitch provides an efficient and economical method to monitor changes in tRNA charging.

1.1.3 Purine-sensing riboswitches

Classes of riboswitches that sense cellular purine concentrations modulate the expression of genes involved in purine biosynthesis (Kim and Breaker, 2008). These RNAs contain a simple ligand-binding domain that can selectively bind purines or modified purines. Currently, four purine-sensing riboswitch groups regulate gene
expression in response to guanine, adenine, 2’-deoxyguanosine (dG), and the modified purine, 7-aminomethyl-7-deazaguanine (preQ₁).

### 1.1.3.1 Guanine and adenine riboswitches

The guanine riboswitch was first identified upstream of the *B. subtilis xpt-pbuX* operon (Mandal *et al.*, 2003). This cluster of genes encodes proteins responsible for purine metabolism and transport (Christiansen *et al.*, 1997). Initial studies of *xpt-pbuX* revealed that expression of the operon is repressed in response to intracellular guanine concentrations. The search for a regulatory protein responsible for the guanine-dependent repression yielded very few candidates. Close inspection of the *xpt-pbuX* leader region uncovered a series of conserved sequence and secondary structural features that were also present upstream of several other genes involved in purine metabolism (Fig. 1.3A) (Mandal *et al.*, 2003). This RNA was later shown to specifically bind guanine and repress expression of the *xpt-pbuX* operon by premature transcription termination. Analysis of the ligand-binding domain indicated that the leader RNA tightly binds guanine with an approximate $K_D$ of 5 nM and discriminates against adenine and guanine-like molecules.

The search for additional guanine riboswitches resulted in the identification of a second class of purine-sensing RNAs that respond to adenine (Fig. 1.3B) (Mandal and Breaker, 2004a). Conserved sequence and secondary features similar to those of the guanine riboswitch were found upstream of genes involved in adenine biosynthesis and
transport. In contrast to the guanine riboswitch, some adenine riboswitches turn on gene expression. Adenine riboswitches that activate gene expression reside upstream of genes involved in purine efflux, such as the \textit{ydhL} gene in \textit{B. subtilis}, which encodes a transporter that removes purines from the cell. Therefore, activation of genes involved in purine efflux is logical as increased purines in the cell necessitate the production of proteins that function to reduce purine concentrations. Other adenine riboswitches repress gene expression through inhibition of translation initiation as seen with the \textit{add} riboswitch from \textit{Clostridium perfringens} (Serganov \textit{et al.}, 2004). Comparison of the guanine and adenine riboswitches uncovered a conserved residue in the core of the molecule (position 74) that differs between the two RNAs. Position 74 contains a conserved U in the adenine riboswitch as compared to a C residue in the guanine riboswitch. This suggested that each leader RNA interacts with the ligand through Watson-Crick base pairing and explains the molecular specificity (Mandal and Breaker, 2004a). A C74U mutation in the guanine riboswitch resulted in an alternate specificity for adenine and caused discrimination against guanine, which supported the theory that ligand specificity is achieved through Watson-Crick base pairing.
Figure 1.3. Secondary structures of representative purine-sensing riboswitches. Secondary structure of each purine riboswitch ligand-binding domain. (A) The guanine-sensing riboswitch from the B. subtilis xpt-pbuX leader region. (B) The adenine riboswitch from Vibrio vulnificus located upstream of the add gene. (C) A representative riboswitch that responds to 2'-deoxyguanosine from Mesoplasma florum. (D) The preQ₁ aptamer found upstream of the queC gene in B. subtilis. The organism and gene under control is shown in the upper left corner near each structure. The nucleotide involved in base pairing with the ligand is shown in blue. The chemical structure of each ligand is depicted in the upper right corner near each aptamer. Modified from Kim and Breaker (2008).
To gain additional insight into the interactions that impart purine specificity, X-ray crystallography was utilized to visualize the global architecture of the RNAs (Batey et al., 2004). Overall, both the guanine and adenine riboswitches adopt similar “tuning fork-like” structures with the purine molecule buried in the core. The binding pocket is formed from residues in the core of the molecule in which the primary determinant for ligand specificity is a Watson-Crick base pair between the cognate ligand and position 74. Single-molecule fluorescence resonance energy transfer (FRET) studies demonstrated that long-range interactions between the terminal loops of the two helices are essential for ligand binding and regulation (Lemay et al., 2006). Analysis of co-transcriptional folding of the B. subtilis pubE riboswitch indicates that the leader RNA exhibits four distinct folding events (Greenleaf et al., 2008). However, only two folding transitions were shown to be adenine-dependent. This study represents one of few analyses that offer an in-depth view of how the effector molecule affects folding of the riboswitch during transcription. Further analysis of this type will offer a detailed view of how riboswitch folding directs the structural transition necessary for regulation.

1.1.3.2 Deoxyguanosine (dG) riboswitch

A variant of the adenine and guanine riboswitches was identified upstream of a ribonucleotide reductase and phosphate transporter in Mesoplasma florum (Kim et al., 2007). These variants are structurally similar to the guanine and adenine-sensing riboswitches but regulate gene expression in response to deoxyguanosine (Fig. 1.3C).
Comparison of the dG RNAs with the guanine and adenine riboswitch sequences identified 39 sequence differences in the *M. florum* RNAs. Most of the nucleotides reside in the core of the molecule. It was postulated that the changes in the core nucleotides result in an alternate RNA specificity despite the structural similarity to the purine riboswitches.

Ribonucleotide reductases catalyze the conversion of ribonucleotides into deoxyribonucleotides, a critical cellular process (Kolberg *et al.*, 2004). Down-regulation of ribonucleotide reductase expression is the result of the presence of dG that promotes a structural change in the riboswitch and results in the formation of an intrinsic terminator upstream of the coding region. Changes in the sequence of the core region, as compared to the purine riboswitches, results in an alternate binding pocket and RNA specificity. Structural analysis of the dG leader RNA indicates that two binding-pocket nucleotides, C58 and C80, are the primary specificity determinants. dG specificity is achieved through modified interactions within the binding pocket where the structure readily accommodates the deoxyribose moiety and creates a binding-pocket specific for dG (Pikovskaya *et al.*, 2011).

1.1.3.3 PreQ$_1$

A class of *cis*-acting RNAs was identified in the leader region of genes involved in the biosynthesis of queuosine. Queuosine is a hyper-modified nucleoside found at the wobble position in the anticodon of some asparagine, aspartate, tyrosine and histidine
tRNAs (Harada and Nishimura, 1972). The riboswitches identified upstream of queuosine biosynthesis genes are specific for the modified purine 7-aminomethly-7-deazaguanine (preQ₁), a queuosine precursor (Roth et al., 2007). The preQ₁ sensitive RNA element is upstream of the queCDEF operon, which encodes the enzymes responsible for queuosine biosynthesis in B. subtilis. Initial phylogenetic analysis of genes involved in queuosine biosynthesis revealed an unusually small aptamer domain that contains as few as 40 nucleotides in some representatives. Genomic searches for putative preQ₁ riboswitches uncovered a widely distributed RNA element with a structure of approximately 106 nucleotides. The queC leader RNA in B. subtilis (Fig. 1.3D), undergoes a structural transition in response to preQ₁. In-line probing of the queC leader from B. subtilis, which monitors cleavage of the 2’OH as a function of ligand concentration, suggests that the RNA binds preQ₁ with an approximate $K_D$ of ~20 nM (Roth et al., 2007). Currently, it is not clear how structural modulation of the aptamer domain affects the regulatory structure as the presence of preQ₁ caused little rearrangement near the putative leader region terminator. Based on phylogeny, the RNA is predicted to regulate at the level of transcription in 22 of the 35 motifs analyzed. The remaining 13 RNAs are predicted to regulate during translation. Further studies are necessary to confirm that regulation of gene expression does occur in the presence of preQ₁. Analysis of the preQ₁ riboswitch within the context of the entire leader RNA will be useful for further characterization of the regulatory function.
1.1.4 Cyclic-di-GMP riboswitches

A recently discovered class of riboswitches that are specific for the di-nucleotide cyclic-di-GMP (c-di-GMP) has been discovered (Sudarsan et al., 2008). The c-di-GMP molecule activates a variety of processes including cell differentiation, biofilm formation, and virulence (Jenal and Malone, 2006; Cotter and Stibitz, 2007). Metagenomic searches first identified the c-di-GMP riboswitch upstream of a group of genes classified as Genes for the Environment, Membranes, and Motility (GEMM) (Weinberg et al., 2007). RNAs with similar motifs were thus termed GEMM elements. Analysis of leader regions upstream of genes that encode the enzymes responsible for the production and degradation of c-di-GMP in Vibrio cholera revealed a region of conserved RNA structure possibly involved in gene expression (Sudarsan et al., 2008). A large number of the known riboswitches repress gene expression in response to the cognate ligand; however, some c-di-GMP leader RNA activate gene expression in response to c-di-GMP. The recognition of c-di-GMP is predicted to differentially regulate gene expression with some riboswitches activate gene expression while others cause repression. The GEMM elements involved in c-di-GMP regulation have been classified based on two distinct structural motifs, designated type 1, and type 2. The two RNA classes have similar architectures with a conserved central helix and GNRA tetraloop; however, the type 1 RNAs contain a tetraloop receptor that is not present in type 2 RNAs. The two RNAs bind c-di-GMP with an affinity in the picomolar range (Smith et al., 2011).
representative riboswitch from the type 2 class was shown to regulate gene expression in *E. coli* in response to c-di-GMP using translational reporter gene fusions (Sudarsan *et al.*, 2008). Mutational analysis of the wild-type translational fusion showed a loss of c-di-GMP-dependent expression upon disruption of the central helix. Compensatory mutations restored expression and suggested that the conserved RNA structure is critical for regulation of expression in response to c-di-GMP.

The small and relatively simple structure of the c-di-GMP riboswitch makes this RNA amenable to x-ray crystallography. Several high-resolution structures have been published that provide the molecular details of the RNA bound to c-di-GMP (Kulshina *et al.*, 2009; Smith *et al.*, 2009; Smith *et al.*, 2011). Similar to the guanine and adenine riboswitches, the c-di-GMP RNA makes standard base pairing interactions with the Watson-Crick face of the second guanine (Gβ) of the dinucleotide (Fig. 1.4). Additional interactions, such as pairing with the first guanine (Gα) and stacking with a universally conserved adenine, construct an intricate network of interactions that contribute to the c-di-GMP binding affinity (Smith *et al.*, 2010c; Smith *et al.*, 2011). Surprisingly, phylogenetic analysis combined with the current structural data has revealed that the most highly conserved residues do not interact directly with the ligand; instead, these nucleotides are critical for base stacking and stabilization of tertiary interactions. This work suggests that the sequence of the RNA is not only critical for recognition of the ligand but also is crucial for formation of the tertiary structure. It is possible that these
interactions are essential for the structural transition required to bind c-di-GMP and modulate gene expression. Variants with mutations in the c-di-GMP binding pocket were analyzed to determine the effect on affinity (Smith et al., 2010a). Binding pocket variants had a negative effect on the c-di-GMP riboswitch but the overall affinity remained in the nanomolar range. Further studies will be necessary to establish the structure-function relationship between distal nucleotides in the central helix and c-di-GMP binding.
Figure 1.4. Structure of the c-di-GMP binding pocket from *V. cholera*. Structure of the binding pocket of the class I *tfoX* c-di-GMP riboswitch bound to c-di-GMP. Bases are numbered. $G_a$ ($G_1$) is involved in Watson-Crick interactions with G19 and A47. The lower guanine $G_b$ ($G_{II}$) is base paired with C92. Dotted lines indicate hydrogen bonds. The c-di-GMP molecule is colored by atom: grey, carbon; blue, nitrogen; and red, oxygen. Adapted from Smith *et al.* (2010b).

1.1.5 The M box riboswitch

Magnesium (Mg$^{2+}$) is the most abundant biological metal in living organisms (Romani and Scarpa, 2000). It plays a pivotal role in a variety of cellular processes including enzyme catalysis and stabilization of RNA structures (Brantl, 2006). Intracellular Mg$^{2+}$ concentrations are critical as too much or too little magnesium can quickly become detrimental to survival. Thus, regulation of genes involved in the transport and uptake of Mg$^{2+}$ is a necessity. One of the best-characterized mechanisms of Mg$^{2+}$ regulation is the PhoP/PhoQ two-component system in *Salmonella enterica* serovar
Typhimurium, one example among a vast number of protein-based regulatory systems. Studies of Mg\textsuperscript{2+} regulation have also identified a leader RNA with the ability to sense Mg\textsuperscript{2+} concentrations in the cell and control the expression of Mg\textsuperscript{2+} transporters (Groisman et al., 2006). Initial reports revealed that mgtA transcripts were differentially expressed in *S. enterica* serovar Typhimurium when grown under various Mg\textsuperscript{2+} conditions (Groisman et al., 2006). A conserved RNA structure was found upstream of the mgtA gene that encodes a PhoP-activated Mg\textsuperscript{2+} transporter. It was proposed that MgtA binds the 5\textsuperscript{'-}UTR of the mgtA transcript when Mg\textsuperscript{2+} conditions are high (Groisman et al., 2006).

Phylogenetic analysis of the *B. subtilis* mgtE gene uncovered a conserved sequence and structural pattern upstream of several Mg\textsuperscript{2+} transport genes (Dann et al., 2007). Closer inspection of the leader RNA structure revealed regions of conserved structure that include transcription terminator and antiterminator elements, suggesting that mgtE is regulated at the transcriptional level. This region of the RNA was designated the M box. *In vivo* analysis of the mgtE leader region fused to lacZ revealed that expression of mgtE-lacZ fusions is induced when Mg\textsuperscript{2+} is limiting while increased Mg\textsuperscript{2+} concentrations caused repression of gene expression. *In vitro* transcription confirmed that regulation of the mgtE transcript occurs via premature transcription termination in the presence of Mg\textsuperscript{2+} and supports the *in vivo* analysis.
X-ray crystal structures of the M box RNA in the presence of Mg\textsuperscript{2+} indicated that six Mg\textsuperscript{2+} ions are bound to the RNA (Dann et al., 2007). The architecture of this RNA is comparable to other small molecule-binding riboswitches in which the core of the molecule scaffolds the binding pocket. Other structural features including base-stacking and conserved structural motifs are utilized to stabilize long-range tertiary interactions. The requirement for six Mg\textsuperscript{2+} ions is unusual as the majority of riboswitches respond to a single ligand. To better understand ligand specificity, additional structural analysis in the presence of manganese (Mn\textsuperscript{2+}) instead of Mg\textsuperscript{2+} uncovered an alternate leader RNA specificity (Ramesh et al., 2011). The results indicated that the leader RNA binds Mn\textsuperscript{2+} with a 4-fold higher affinity than Mg\textsuperscript{2+}; this suggests that the M box riboswitch may not be specific for magnesium, but may sense a variety of divalent cations. A better understanding of the specificity of the M box riboswitch will be of particular interest given that the leader RNA responds to a pervasive subset of molecules.

1.1.6 The \textit{glmS} ribozyme

The discovery of small molecule-binding riboswitches has prompted a number of large-scale bioinformatics studies focused on the identification of previously uncharacterized riboswitches. One such study identified a conserved leader region upstream of the \textit{glmS} gene (Barrick et al., 2004). The \textit{glmS} gene in \textit{B. subtilis} encodes glutamine-fructose-6-phosphate amidotransferase (Milewski, 2002). This enzyme is responsible for the conversion of fructose-6-phosphate and glutamine into glucosamine-
6-phosphate (GlcN6P), the first committed step of cell wall biosynthesis. Phylogenetic analysis of the B. subtilis glmS leader revealed an RNA that includes three stem loops with a pseudoknot, and a minimally conserved core; however, this analysis did not identify the expression platform (Barrick et al., 2004; Winkler et al., 2004).

In vitro analysis of the glmS leader RNA uncovered an alternative mechanism for regulation of the glmS gene (Winkler et al., 2004). The glmS leader RNA down-regulates gene expression in response to GlcN6P by endonucleolytic cleavage of the glmS transcript. When GlcN6P concentrations are high, GlcN6P binds the leader RNA and promotes self-cleavage of the RNA upstream of the glmS coding region; cleavage of the transcript results in repression of glmS expression. A significant reduction in cleavage was observed when the glmS ribozyme was transcribed in the presence of GlcN6P analogs, as compared to Glc6P. This suggests that the ribozyme selectively binds GlcN6P, which specifically promotes cleavage of the glmS transcript. Mutation of conserved residues in the glmS ribozyme resulted in derepression of gene expression in the presence of GlcN6P. The glmS cleavage reaction occurs via nucleophilic attack of the adjacent phosphorus. Studies of the substrate requirements for cleavage indicate that GlcN6P acts as a coenzyme and facilitates the chemical cleavage. It was shown that the amine group of the GlcN6P molecule is essential; thus, Glc6P and other analogs that lack primary amines were unable to promote efficient cleavage (McCarthy et al., 2005). The cleavage products are transcripts that contain a 5’-OH, which is a substrate for
riboinuclease (RNase) J1 (Collins et al., 2007). In vivo analysis confirms that the truncated \textit{glmS} transcript is targeted for degradation. Cleavage of the \textit{glmS} transcript in response to GlcN6P represents a unique mechanism that employs both sensory and catalytic RNA functions to control gene expression. RNA-based regulation of \textit{glmS} in Gram-positive organisms is an alternative mechanism to regulate gene expression by targeting the mRNA for degradation.

1.1.7 Coenzyme-sensing riboswitches

RNAs of this class control the expression of genes responsible for the synthesis of essential metabolic co-factors, vitamin B\textsubscript{1} (TPP), vitamin B\textsubscript{2} (FMN), vitamin B\textsubscript{12} (AdoCbl). One of the best-studied riboswitch RNAs regulates gene expression in response to thiamine pyrophosphate (TPP). The TPP-responsive RNAs represent some of the most widespread riboswitches known to exist in all three domains of life.

1.1.7.1 THI box riboswitch (TPP, vitamin B\textsubscript{1})

Initial phylogenetic studies of the 211-base leader RNA upstream of the \textit{thiCOGE} operon in \textit{Rhizobium etli} uncovered a conserved 38 nt region that was designated the THI box (Miranda-Rios et al., 2001). The THI box was shown to be responsible for the repression of thiamin biosynthesis in the presence of thiamin. This conserved sequence is housed in the ligand-binding domain of a riboswitch that regulates gene expression in response to TPP. Transcriptional \textit{thiC-lacZ} fusions from \textit{R. etli} demonstrated that growth
in the presence of thiamin caused a 19-fold decrease in gene expression as compared to expression when thiamin was absent (Miranda-Rios et al., 2001). Deletion of the THI box caused loss of thiamin-dependent repression. This work demonstrated that thiC expression is regulated at the transcriptional level in which a transcriptional terminator forms in the presence of thiamin and causes repression of gene expression. In contrast to the R. etli thiC riboswitch, the E. coli thiM riboswitch contains a SD sequestering helix that prevents translation initiation. Structural probing of the E. coli thiM SD indicated that the presence of TPP caused a structural change in the RNA that resulted in the occlusion of the SD (Winkler et al., 2002). Further analysis of the SD region confirmed that TPP causes repression of gene expression in E. coli through sequestration of the SD (Ontiveros-Palacios et al., 2008). TPP riboswitches are widely distributed in bacteria and have been identified in eukaryotes. The filamentous fungus Aspergillus oryzae contains a TPP riboswitch that regulates gene expression at the level of mRNA splicing. In this example, a conserved THI box RNA is located in the 5’-UTR of the A. oryzae thiA gene (Kubodera et al., 2003). The THI box riboswitch is adjacent to a critical splice site in an intron upstream of the thiA coding sequence, and the presence of TPP promotes a structural rearrangement that inhibits the availability of the splice site for mRNA processing.

Several crystal structures of the THI box RNA bound to TPP are available (Edwards and Ferre-D'Amare, 2006; Serganov et al., 2006; Thore et al., 2006). X-ray
crystallography of the *E. coli thiM* aptamer domain illustrates the complex global RNA structure of the molecule. The THI box RNA contains three compact helices connected by a three-way junction (Serganov *et al.*, 2006). Two of the helices are positioned in a parallel conformation that becomes the scaffolding for the TPP binding pocket (Fig. 1.5). The TPP molecule is bound in an extended conformation that forms an inter-domain bridge across the two helices (Fig. 1.5).
Figure 1.5. X-ray crystal structure of the TPP bound aptamer domain. Crystal structure of the *E. coli* thiM aptamer domain bound to TPP. Stick-and-ribbon representation of the front (A) and top (B) views of the aptamer. Two helices are aligned to form the TPP binding pocket, which is enclosed by the P1 helix. The RNA completely encapsulates the TPP molecule (red) which is bound in an extended conformation. Colors and labels denote RNA stems (P), loops (L), and junctions. Modified from Serganov *et al.* (2006).

It was suggested that the TPP molecule stabilizes the global RNA structure by bridging the two helices (Serganov *et al.*, 2006). TPP recognition occurs through extensive hydrogen bonding interactions between the RNA and the 4-amino-5-hydroxymethyl-2-methylpyrimidine ring as it is intercalated between two adjacent bases.
(Edwards and Ferre-D'Amare, 2006). The phosphate region of the molecule is recognized with the support of two Mg$^{2+}$ ions that aid in shielding the negative charge. The conformation of the TPP molecule allows for recognition of additional functional groups and increases the RNA specificity for TPP. This is predicted to be one mechanism that the RNA uses for discrimination between thiamin monophosphate, thiamin, and TPP (Serganov et al., 2006).

The TPP riboswitch exhibits exquisite specificity for TPP and discriminates against closely related analogs. This is a common trait for small-molecule binding riboswitches. The TPP riboswitch is unique in that it is the only riboswitch found in all three domains of life. It is possible that additional riboswitches are more widespread and have yet to be discovered. It is interesting that most known riboswitches control expression via transcription termination or translational initiation. The identification of the TPP riboswitch in A. oryzae was the first indication that riboswitches can also control mRNA splicing. The TPP riboswitch employs alternative methods to control gene expression; currently it is unclear which mechanisms are more effective. Analysis of the evolutionary trends of regulation by the TPP riboswitch may offer insight into why it may be more advantageous to regulate during transcription for some organisms as compared to translation.
1.1.7.2 Vitamin B$_2$, flavin mononucleotide (FMN)

A second class of vitamin-sensing RNAs was identified and was designated RFN elements (Winkler et al., 2002). These RNAs regulate genes involved in the synthesis and transport of riboflavin. In *B. subtilis*, the *ribDEAT* (*ribD*) operon is responsible for the synthesis of riboflavin. Transcription of the 5’ leader region of the *ribD* operon in the presence of FMN, the biologically active form of riboflavin, results in transcription termination (Winkler et al., 2002). In contrast to the *ribD* leader RNA, the *ribU* leader RNA is predicted to regulate the synthesis of a riboflavin transport protein at the level of translation through modulation of the accessibility of the SD. The predicted secondary structure of the RFN element revealed a large RNA with six helices connected by a conserved core, similar to other known riboswitches. Further analysis of the aptamer region confirmed that the RFN element binds FMN directly and causes structural rearrangement of the RNA. Ligand recognition by this RNA was of particular interest considering that FMN has a number of biologically relevant analogs. Riboflavin is an FMN precursor that lacks a phosphate group. Initial analysis of the concentration of FMN required to promote half-maximal cleavage during in-line probing analysis estimated that the *B. subtilis* RFN element binds FMN with a $K_D$ of ~ 5 nM. Analysis of the riboflavin interaction with the RFN element indicated that the ligand was bound with a $K_D$ of 3 μM. The increased $K_D$ for FMN demonstrates that the RFN element can differentiate between two chemically similar molecules, possibly through recognition of the additional phosphate group.
X-ray crystallography of the *impX* RFN element, from *Fusobacterium nucleatum*, bound to FMN provided additional molecular details of FMN recognition (Serganov *et al.*, 2009b). Helical stacking is a common structural mechanism utilized by riboswitch RNAs for scaffolding and organization of the ligand-binding domain. However, the RFN element does not use coaxial stacking for organization of the global structure. Instead, the RNA adopts a butterfly-like fold that is stapled together by several long-range interactions between the six helices (Fig. 1.6A).

Crucial tertiary interactions between the loops and helices act as molecular staples that stabilize the RNA structure (Fig. 1.6A) (Serganov *et al.*, 2009b). Similar to other riboswitch structures, the FMN binding pocket is centrally located within the core of the molecule and surrounded by six helices (Fig. 1.6B). When FMN is bound, the RNA completely engulfs the molecule and directly interacts with the isoalloxazine rings. Specific interactions with the Mg$^{2+}$ ions also stabilize the binding pocket and provide counter-ion shielding for the negatively charged phosphate moiety of FMN.

The structures of the RFN element bound to TPP analogs riboflavin (lacks the phosphate group) and roseoflavin (lacks phosphate and has additional dimethylamine) illustrate that alternative RNA-ligand interactions are concentrated near the ribityl moiety of each analog (Fig. 1.7). Although riboflavin lacks the phosphate group, adjustment of the P4 helix (Fig. 1.6A) and positioning of the binding pocket nucleotides (Fig. 1.7A) provide a RFN element that readily accommodates the additional dimethylamino group.
The organization of the RFN element binding pocket when bound to roseoflavin was similar to the structure when bound to riboflavin (Fig. 1.7B), although minor changes were observed.

**Figure 1.6. Structural models of the RFN element.** (A) Secondary structural model of RFN element. Colors with coordinating letters and numbers indicate loops (L), helices (P), and junctions (J). Black arrows indicate positions of long-range interactions and tertiary interactions. (B) Tertiary structural model of RFN element from *F. nucleatum* in complex with FMN. The coloring is the same as shown in panel A. FMN shown in red. Adapted from Serganov *et al.* (2009).
Figure 1.7. Structures of the *F. nucleatum* impX RFN element bound to FMN and analogs. (A) 2.9 Å crystal structure of the *impX* leader RNA binding pocket bound to riboflavin (blue and green) superimposed over the binding pocket bound to FMN (grey). (B) Structure of the binding pocket bound to roseoflavin (pink) superimposed over the riboflavin bound pocket (blue and green). Adapted from Serganov *et al.* (2009).

Recent work on riboswitches has focused on the unliganded structure of the RNAs (Liberman and Wedekind, 2011). Structural mapping and X-ray crystallography studies by Vicens *et al.* (2011) uncovered the unliganded structures of the *F. nucleatum* and *B. subtilis* RFN elements. The results suggest that large conformational changes in the aptamer domain are due to the Mg$^{2+}$ concentration as compared to minimal structural changes that were observed in the presence of FMN (Vicens *et al.*, 2011). Current riboswitch models predict that conformational changes in RNA structure are the result of ligand binding (Winkler *et al.*, 2002). However, additional X-ray crystallography data supports initial findings that the FMN riboswitch is folded into a bound-like state in the
absence of the ligand (Vicens et al., 2011). Data from selective 2’-hydroxyl acylation analyzed by primer extension (SHAPE) experiments, which monitors the reactivity of the 2’-OH regions with N-methylisatoic anhydride (NMIA), provided a description of the RFN element in the presence and absence of FMN. The structural mapping reveals that four of the six single-stranded junctions undergo changes due to FMN binding. As with many other riboswitches, the distinct structural transitions that occur upon signal recognition remain elusive. The structural complexity of the RFN element and organization of the global RNA structure makes elucidation of the structural changes that occur in response to FMN particularly difficult. However, resolution of the structural changes that occur upon ligand binding will be particularly interesting, as it will provide the critical details of what makes the riboswitch switch.

1.1.7.3 Vitamin B\textsubscript{12}, 5’-deoxy-5’-adenosylcobalamin (AdoCbl)

At least 200 B\textsubscript{12} elements have been identified representatives in both Gram-positive and Gram-negative organisms (Vitreschak et al., 2003). The B\textsubscript{12} riboswitch contains at least 10 base-paired helices and is one of the largest and most complex riboswitch structures (Nahvi et al., 2002). This highly conserved element and contains at least 57 nucleotides with >90% conservation (Nahvi et al., 2004). In E. coli and S. typhimurium, the btuB gene encodes an outer membrane cobalamin transport protein. The 202-nt btuB riboswitch in E. coli directly binds vitamin B\textsubscript{12} with high affinity and in response undergoes significant structural rearrangement. The btuB riboswitch serves as a
typical genetic “off” switch that represses gene expression when AdoCbl is present. Regulation of *btuB* occurs at the level of translation initiation through formation of a hairpin structure that sequesters the SD. Disruption of the hairpin structure that sequesters the SD causes a loss of AdoCbl-dependent repression.

Structural characterization of riboswitch-ligand complexes has been instrumental in our current understanding of ligand recognition. In many cases, the ligand is completely engulfed by the RNA through a complex network of interactions. Currently, the details of AdoCbl recognition are unavailable; however, several biochemical methods have been utilized to examine the AdoCbl specificity. Detailed analysis of B12 binding revealed that the corrin ring is the primary determinant for ligand recognition (Gallo *et al.*, 2008). Recent identification of the B12 riboswitch upstream of the *eut* operon in *Enterococcus faecalis* has revealed that the riboswitch can also regulate by transcription termination (Baker and Perego, 2011). The leader region of the *eut* operon in *E. faecalis* was analyzed both *in vitro* and *in vivo* and found to terminate transcription in the presence of AdoCbl. Additional studies of the *eut* riboswitch revealed that the leader RNA is located between two promoters and may potentially be involved in additional regulatory processes related to ethanolamine utilization.
1.1.8. S box riboswitch

1.1.8.1. S box (SAM-I)

One of the first reports of a small-molecule-binding riboswitch revealed the S-adenosylmethionine (SAM) responsive leader RNA termed the S box. S box RNAs were initially found in the genomes of low-G+C Gram-positive bacteria (Grundy and Henkin, 1998). Since the initial discovery, the S box has been one of the most extensively studied riboswitches. Analysis of 11 leader regions of genes involved in methionine or cysteine metabolism revealed areas of sequence and structural conservation (Grundy and Henkin, 1998). Similar features were found upstream of genes in a number of organisms. The conservation of these features suggested a widespread regulatory mechanism that was designated the S box regulon. One feature of the 11 conserved structural elements was an intrinsic transcriptional terminator, which suggested that regulation might occur by premature transcription termination. A structural model of the S box leader RNA was proposed based on the phylogenetic analysis of 18 leader regions. The model predicted that the leader RNA could form two alternate structures. One conformation included an intrinsic terminator that when formed would terminate transcription prior to RNAP reaching the coding region. In contrast, the alternate conformation consisted of an antiterminator element that when formed would allow transcription to continue. One key feature of the two structures was that the terminator and antiterminator helices were mutually exclusive as the same region of RNA sequence was required for the formation
of each independent structure (Fig. 1.8). In addition to the two regulatory elements, the predicted structure also contained an anti-antiterminator element that would prevent terminator formation by the sequestration of sequences required by the antiterminator (Fig. 1.8). A conserved structural motif termed the GA (kink-turn) motif was also identified in the S box leader region (Winkler et al., 2001). This motif is conserved in the T box leader RNA and it was suggested that it might be instrumental in the formation of the tertiary structure. The degree of conservation and the complexity of the predicted structures suggested that the RNA could act as a binding site for an effector molecule and regulate gene expression through structural modulation (Fig. 1.8).
Figure 1.8. S box riboswitch. Shown is a representative S box (SAM-I) riboswitch that regulates at the level of transcription. The aptamer (AAT) region consists of four helices (P1-P4). In the absence of SAM, the antiterminator (AT) forms and transcription of the downstream gene continues (black arrow). SAM (*) binding by the AAT stabilizes the P1 helix and prevents formation of the AT. SAM binding also sequesters nucleotides required for the formation of the AT and allows terminator (T) formation. Red and blue lines denote mutually exclusive sequences. A circled asterisk represents the SAM molecule and the dotted line indicates a long-range tertiary interaction responsible for formation of SAM binding pocket. Adapted from Smith et al. (2010a).

To confirm the predicted regulatory mechanism, the yitJ, ykrT, and ykrS leader sequences from B. subtilis, all members of the S box regulon, were fused to lacZ and expression was monitored in response to methionine limitation (Grundy and Henkin,
Expression of lacZ fusions was reduced when methionine was high as compared to expression during non-limiting conditions. Mutations in the conserved sequence and structural features (other than disruption of the antiterminator) resulted in constitutive expression of the reporter gene. These results indicated that the leader RNA is responsible for the genetic control and that the structure is linked to the regulatory function.

Further genomic searches uncovered additional S box leader RNAs (McDaniel et al., 2003). In vitro analysis of the yitJ leader RNA from B. subtilis provided the first evidence that the presence of SAM causes premature transcription termination (Epshtein et al., 2003; McDaniel et al., 2003; Winkler et al., 2003). Expression of the yitJ gene was shown to be regulated by premature termination of transcription in the presence of SAM both in vivo and in vitro (McDaniel et al., 2003; Winkler et al., 2003). Other potential effectors were tested in vitro and none was able to promote termination similar to SAM. Mutation of conserved residues in the yitJ leader caused a reduction of in vitro termination and resulted in constitutive expression (McDaniel et al., 2003). The S box RNA binds SAM with high affinity and specificity. The binding affinity for the SAM-RNA complex was estimated in the low micromolar range for the yitJ RNA (McDaniel et al., 2003; Winkler et al., 2003).

Biochemical analysis of the S box leader structural transition demonstrated that formation of the SAM-dependent transcription terminator also requires the stabilization
of an additional anti-antiterminator structure (Fig. 1.8) (McDaniel et al., 2003). Crystal structure analyses of the anti-antiterminator regions of S box leader RNAs from *Thermoanaerobacter tengcongensis* and *B. subtilis* are consistent with predicted tertiary interactions such as the presence of a pseudoknot and GA motif (McDaniel et al., 2005; Montange and Batey, 2006; Lu et al., 2010). The formation of both RNA motifs is essential to the function of the leader RNA (Heppell and Lafontaine, 2008; Heppell et al., 2011). The SAM binding pocket is formed in the core of the molecule and the global structure of the RNA is anchored by essential tertiary interactions between covarying regions and conserved RNA motifs (Montange and Batey, 2006). The high-resolution structure of the *yitJ* S box RNA from *B. subtilis* demonstrates that binding pockets of leader RNAs from two different organisms are nearly identical although they differ in the structure of peripheral elements (Lu et al., 2010). Biochemical analysis by SHAPE indicates that the leader RNA forms a pre-binding conformation that differs from the structure observed when SAM is bound. The chemical probing further revealed that the conformational switch that occurs in the presence of SAM is primarily directed by the alternative folding of sequences involved in formation of the expression platform. The binding pocket is optimized to bind SAM and completely engulfs the molecule. The sulfur and methyl group of the methionine molecule directly interact with the RNA and provide a mechanism for the RNA to discriminate against similar molecules such as SAH (Montange and Batey, 2006). Analysis of 11 S box leader RNAs revealed a natural variability in SAM sensitivity (Tomsic et al., 2008). Comparison of the expression
patterns of S box transcriptional units suggests that genes that are directly involved in methionine biosynthesis are more tightly regulated as opposed to genes involved in methionine transport and uptake. The variable SAM sensitivity observed in the S box leader RNAs suggests that even riboswitches that respond to the same molecule can be titrated to specific ligand concentrations.

1.1.8.2 SAM-II

A second class of SAM-responsive leader RNAs, structurally distinct from the S box RNAs, are found in α-proteobacterial genomes upstream of methionine biosynthesis genes (Corbino et al., 2005). SAM-II RNAs are predicted to regulate expression during either transcription or translation. The conserved sequence and secondary structure is fewer than 70 nt in length and predicted to form a pseudoknot-like structure (Corbino et al., 2005). Several structural analyses have been performed that indicate that the RNA adopts a simple H-type pseudoknot in which base triple interactions with the SAM molecule are responsible for recognition (Gilbert et al., 2008; Kelley and Hamelberg, 2010). Crystal structural analysis indicates that the SAM-II riboswitch uses a strategy similar to that of the S box RNAs to recognize SAM. One example is the formation of a base triple between binding pocket nucleotides and the adenine moiety, which is similar to the S box. Recognition of the positively charged sulfur molecule occurs through interactions with the carbonyl oxygen of conserved U residues. This is similar to the interactions between sulfur group of SAM and a conserved A-U pair found in the SAM
binding pocket. Although SAM is recognized directly by the SAM-I riboswitch, analogs with modifications at the thioether group such as sulfone, sulfoxide, and aza derivatives were also bound by the RNA, albeit with reduced affinity (Lim et al., 2006).

1.1.8. 3 S\(_{\text{MK}}\) box (SAM-III)

SAM-sensitive riboswitches that regulate the expression of SAM synthetase genes also reside in the genome of members of the Lactobacillus order (Fuchs et al., 2006). These RNAs were designated S\(_{\text{MK}}\) box riboswitches since they were identified upstream of \textit{metK} genes. Phylogenetic analysis uncovered a helix overlapping the SD region, which suggested that regulation might occur during translation initiation. To investigate the role of the S\(_{\text{MK}}\) box RNA in translation initiation, the \textit{E. faecalis} met\textit{K} leader RNA was fused to a \textit{lacZ} reporter and integrated into the \textit{B. subtilis} chromosome. Under high SAM conditions, gene expression was repressed, and expression increased after a reduction in SAM pools. Subsequent biochemical analysis demonstrated that the leader RNA forms a structure in the presence of SAM that sequesters the SD sequence and prevents binding of the ribosome (Fuchs \textit{et al.}, 2006; 2007). Occlusion of the ribosome was specific as 30S ribosomes were able to bind and initiate translation in the presence of SAH (Fuchs \textit{et al.}, 2006). This work demonstrates that the S\(_{\text{MK}}\) box leader RNA specifically regulates expression in response to SAM and discriminates against similar molecules.
Figure 1.9. Structure of $S_{MK}$ box riboswitch. (A) Predicted secondary structure of the $S_{MK}$ box riboswitch from phylogenetic analysis. Black letters indicate residues that are 100% conserved; blue 50-85% conserved; R, G or A; W, A or U; Y, C or U. Orange dotted lines represent hypervariable regions. Solid ladders indicate regions of conserved secondary structure. Gray shaded area denotes SD sequence and pink shaded area is AUG start codon. (B) Cartoon representation of the $S_{MK}$ box riboswitch crystal structure. SAM is shown in overlapping CPK (magenta) and surface (silver) representations. Helices P1-P4 are colored in cyan, green, silver, and yellow, respectively. Adapted from Lu et al. (2008).

Structural analysis of the leader RNA confirmed secondary structural models and provided evidence of a riboswitch with a combined aptamer and expression platform similar to the SAM-II S box RNA (Lu et al., 2008). Most riboswitches contain a dual domain architecture in which the aptamer and expression platform are independent.
structures. However, the S$_{MK}$ box RNA adopts an inverted Y shape that binds SAM and sequesters the SD (Fig. 1.9). The SD sequence contacts SAM directly to form the SAM binding pocket (Lu et al., 2008). Additionally, further studies of the E. faecalis metK riboswitch demonstrate that the S$_{MK}$ leader RNA is a reversible riboswitch with the capacity to participate in multiple SAM-dependent regulatory events during the lifetime of the RNA (Smith et al., 2010a). Chemical probing methods were utilized to uncover details of the folding landscape of the S$_{MK}$ box RNA (Lu et al., 2011). The results of this work predict a folding landscape for the RNA that illustrates the importance of the sequence-function relationship. Mutation of the RNA in hypervariable regions resulted in an altered folding landscape. This highlights the importance of structural elements that are not in close proximity to the ligand-binding site and demonstrates that although binding is important there are other key RNA features that contribute to the function.

1.1.8.4 S box variants

Two groups of RNAs have been identified that contain SAM-binding recognition and characteristics similar to the S box leader RNAs. However, the global RNA structure of these RNAs differs from the known S box leader RNAs (Weinberg et al., 2007). The first group, designated SAM-IV is predicted to regulate expression during translation (Weinberg et al., 2008). Although, in-line probing data indicate that a 132-nt Streptomyces coelicolor leader RNA undergoes structural modulation in response to
SAM. Currently, little biochemical evidence is available to confirm the regulatory mechanism.

The second group of variants (SAM-V) are found only marine $\alpha$-proteobacteria (Poiata et al., 2009). These RNAs are similar to the SAM-II class of riboswitches and are located in the intergenic regions of a number of methionine biosynthesis genes. The $\text{met}Y$ gene in Candidatus Pelagibacter ubique encodes $O$-acetylhomoserine (thiol)-lyase, an enzyme that produces L-methionine from L-homocysteine. In the presence of SAM, the $\text{met}Y$ RNA undergoes significant structural modulation, specifically binds SAM, and strongly discriminates against SAM analogs. The sequence of the RNA suggests that SAM-dependent regulation may occur during translation by occlusion of the RBS. Additionally, this variant also exists in a tandem arrangement with the SAM-II leader RNA. This tandem riboswitch architecture is worth further investigation to determine if this variant is directly involved in the regulation of gene expression when in tandem, or if they are only present as SAM-sensing domains to support the function of the SAM-II riboswitch.

1.1.8.5 $S$-adenosylhomocysteine (SAH) riboswitch

Highly conserved bacterial RNA elements, designated SAH elements, were discovered in $\beta$ and $\gamma$-proteobacteria upstream of genes involved in $S$-adenosylhomocysteine (SAH) recycling. SAH is a by-product of the SAM-dependent methyltransferase reactions (Wang et al., 2008). Genomic searches for additional
riboswitches identified at least 68 conserved RNA elements upstream of genes involved in SAH recycling. These genes are involved in the degradation of SAH to adenosine and homocysteine. Homocysteine is converted to methionine, the substrate for SAM synthesis. The aptamer region of the achY leader RNA from *Pseudomonas syringae* binds tightly to SAH and discriminates against SAM by 1000-fold. Analysis of the molecular recognition features indicated that the leader RNA utilizes every functional group for SAH recognition. achY encodes SAH hydrolase, which catalyzes the degradation of SAH into adenine and homocysteine. In vivo analysis of cells harboring achY translational fusions demonstrated that the SAH element is responsible for the induction of gene expression when SAH concentrations are high. The activation of achY expression is similar to the rationale of regulation by the adenine responsive riboswitch, wherein the presence of SAH promotes activation of genes involved in the degradation of SAH. Regulation of the achY gene occurs during translation initiation. In the absence of SAH, there is low achY expression due to the sequestration of the SD within the SAH element. However, an increase in SAH concentrations causes the RNA to undergo a structural change that releases the SD and makes it available for translation. Mutation of conserved residues inhibited SAH-dependent induction and resulted in repression of gene expression. Additional mutations that destabilized the structure of the leader region resulted in constitutive expression of the reporter gene. The high gene expression was suggested to be the result of the mutations, which caused a loss of regulatory function due to misfolding of the RNA. Although the primary mechanism of regulation of SAH
catabolism is at the translational level, the metH gene from Dechloromonas aromatica is predicted to regulate gene expression by transcription termination.

Recent crystal structure analysis of the transcriptionally regulated achY leader from Ralstonia solanacearum reveals a complex RNA topology and demonstrates the molecular basis for SAH recognition (Edwards et al., 2010). The structure of the RNA is organized into an “LL type” pseudoknot that forms the SAH binding pocket in the core of the RNA. SAH recognition occurs through a series of hydrogen bonds with the adenine ring and methionine moiety of the molecule (Fig. 1.10). It is postulated that the interaction between the sulfur group and RNA backbone helps to distinguish between SAH and the structurally similar SAM.
Figure 1.10. **Crystal structure of SAH binding pocket.** Crystal structure of the SAH binding pocket from the *R. solanacearum ahcY* riboswitch. The binding pocket is composed of highly conserved nucleotides hydrogen bonded to the main chain of SAH. SAH is depicted in salmon, base interactions are illustrated using the notation of Leontis and Westhof (2001). Dashed lines represent hydrogen bonds and the double arrow denotes the distance between the sulfur of SAH and the O4’ of A29. Adapted from Edwards *et al.* (2010).

Superimposition of SAM in a configuration similar to SAH as seen in the crystal structure suggests that the presence of the methyl group, which distinguishes SAM from SAH, would result in a steric clash with the RNA backbone (Edwards *et al.*, 2010). SAM is not the only biologically relevant SAH analog, thus 15 other SAH analogs were assayed to determine the specificity of the SAH element (Wang *et al.*, 2008). The RNA bound several molecules; however, none bound as tightly as SAH. To explore the basis of SAH specificity, additional adenine-bearing molecules, such as ATP and NAD⁺, were
tested in a binding assay (Edwards et al., 2010). The SAH aptamer could bind both derivatives with high affinity. It is interesting that other adenine-bearing molecules are able to bind the SAH element with an affinity similar to that of the cognate effector. Although these compounds may bind the RNA, it is unclear whether they also modulate gene expression. Functional analysis of the SAH element with regard to ligand specificity will provide useful information on the true effector and support the characterization of the SAH element.

1.1.9 Amino acid-sensing riboswitches

1.1.9.1 Glycine

Glycine-sensitive riboswitches serve as genetic “ON” switches, and activate gene expression in response to the presence of glycine. The glycine riboswitch can be found in both Gram-negative and Gram-positive organisms; however, they are more prevalent in Proteobacteria, Firmicutes, Fusobacteria and Actinobacteria (Serganov and Patel, 2009a). The best-studied glycine riboswitch is found upstream of the gcvt operon in B. subtilis. The gcvt operon encodes proteins involved in glycine cleavage and consumption (Mandal et al., 2004b). In vivo analysis of the B. subtilis gcvt operon demonstrated that an increase in gene expression occurs in response to glycine.

The gcvt riboswitch has an unusual topology as compared to other known riboswitches. Most riboswitches are composed of two domains, the aptamer domain and
expression platform. However, the glycine ligand-binding domain contains two aptamers upstream of the expression platform (Fig. 1.11) (Mandal et al., 2004b). In-line probing of the dual aptamer region suggests that each domain undergoes a structural rearrangement in response to glycine. This organization allows genetic control to occur in response to two glycine molecules instead of one. It was suggested that the requirement for two glycine molecules, as compared to one effector for other riboswitches, results in regulation that occurs in response to subtle changes in cellular glycine concentrations.

![Figure 1.11. Glycine riboswitch aptamers.](image)

Secondary structure representation of the glycine aptamer region. Aptamers are labeled I and II. Junctions (J), stems (P), and loops (L) are numbered. Red nucleotides are conserved in > 95% and blue nucleotides conserved in >75% of sequences analyzed. Solid lines indicate variable regions. Adapted from Huang et al. (2010).
Equilibrium dialysis of the tandem aptamers indicated that glycine is bound at a 1:1 ratio such that each aptamer binds one glycine molecule (Mandal et al., 2004b). Further stoichiometric studies indicated that glycine binding to aptamer I positively influences the glycine affinity of aptamer II. Glycine binding was found to be cooperative as mutations in aptamer I negatively affected the function of aptamer II. Structural probing of a structurally similar glycine riboswitch from *F. nucleatum* suggested that the P1 helix of aptamer I is involved in tertiary interactions with aptamer II and may explain the observed cooperativity (Kwon and Strobel, 2008). Disruption of helix 1 in aptamer I results in lower affinity binding by aptamer II. Crystal structural analysis of aptamer II from *F. nucleatum* revealed an extensive molecular network around the bound glycine molecule (Huang et al., 2010; Butler et al., 2011). The glycine-binding pocket is adjacent to the single stranded core and contains two Mg$^{2+}$ ions. The intramolecular interactions between the two ligand-binding domains stabilize this tandem architecture through A-minor and stacking interactions that support the formation of a pseudosymmetric dimer.

In genetic regulation by riboswitches, the only requirement is a functional RNA that will regulate gene expression in response to intracellular conditions. The glycine riboswitch is an example of how even simple systems can adjust the regulatory response based on the specific needs within the cell. The glycine riboswitch illustrates that duplication of the aptamer domain can alter the regulatory response and allow regulation
to occur in response to discrete changes in glycine concentrations. It is possible that other riboswitch characteristics can be altered for specific regulatory requirements.

1.1.9.2 Glutamine

Computer-assisted genomic and metagenomic searches were essential in the identification of a potential glutamine responsive riboswitch (Ames and Breaker, 2011). Phylogenetic analysis of conserved RNA motifs upstream of genes involved in nitrogen metabolism revealed a series of conserved sequence and structural elements. The two motifs are approximately 60 nucleotides in length, with three helical regions radiating from a centrally conserved core. Comparison of the two motifs identified RNAs with similar features that were designated the glnA motif and Downstream Peptide (DP) motif. The glnA motif can be found in dual and triple aptamer arrangements in the 5’-UTR of genes involved in nitrogen recycling. The distribution of the motifs within the cyanobacterial genome revealed that the motifs are primarily upstream of genes involved in nitrogen metabolism such as ammonium transporters or glutamine and glutamate synthases. In-line probing of the glnA RNA motif demonstrated that the presence of glutamine and not glutamate caused a structural transition within the RNA (Ames and Breaker, 2011). The presence of multiple aptamer domains upstream of the regulated genes prompted binding analysis to determine if the domains bound glutamine cooperatively. However, no cooperative binding was observed. The lack of cooperative binding presents a potential problem relative to the function of the riboswitch, as the
phylogenetic analysis did not identify a regulatory structure adjacent to each aptamer domain. Although transcription terminators and SD sequestering helices are located downstream of the \textit{glnA} motif, the amount of sequence that separates the structures makes the mechanism of regulatory control difficult to ascertain. This is in contrast to other riboswitch structures in which the regulatory structure is immediately adjacent to the aptamer domain to allow the alternative folding of mutually exclusive structures, a common mechanism that riboswitches employ to regulate gene expression. Future studies focused on how signal recognition by the aptamer domain is communicated to the regulatory structure will be of particular interest and can be applied to our current understanding of riboswitch function.

1.1.9.2 The L box

Lysine is a requirement for a large number of biochemical processes in most organisms and is synthesized \textit{de novo} in bacteria and plants. In most bacteria, lysine is synthesized via the diaminopimelate (DAP) pathway (Patte, 1994; Belitsky, 2002). However, certain bacterial species utilize the \(\alpha\)-aminoadipate pathway for lysine synthesis (Nishida \textit{et al}., 1999). The lysine biosynthetic pathway is required for protein synthesis, cell wall biosynthesis, sporulation, the stress response, and bacterial virulence (Bugg and Brandish, 1994; Torres \textit{et al}., 2007). The first committed step in the pathway is catalyzed by aspartokinase II (AKII), the product of the \textit{lysC} gene (Patte, 1994). AKII is responsible for the phosphorylation of aspartate, a precursor of lysine, methionine, and
threonine biosynthesis (Belitsky, 2002; Serganov and Patel, 2009a). Other essential macromolecules produced in this pathway are 2,6-dipicolinate, a major component of bacterial endospores, and DAP, a key cell wall constituent in bacteria. Early studies of lysine biosynthesis by Lu et al. (1992) demonstrated that transcription of the *B. subtilis* *lysC* gene is repressed 100-fold in the presence of lysine. Mutations in the *lysC* leader region caused derepression of expression and resulted in resistance to the toxic lysine analog aminoethylcysteine (AEC) (Lu et al., 1992; Patte et al., 1998). AEC differs from lysine by a sulfur substitution of a methyl group in the carbon side chain (Girolamo et al., 1982). Further investigation of the mechanism of repression revealed that lysine controls gene expression by premature transcription termination (Kochhar and Paulus, 1996). At the time, the mechanism for transcriptional attenuation was not clear.

Phylogenetic analysis of lysine biosynthesis genes from a number of organisms revealed a set of conserved leader RNA features upstream of a transcriptional terminator; this region was designated the L box (Grundy et al., 2003; Rodionov et al., 2003; Sudarsan et al., 2003). L box RNAs are often found in the Gram-positive Firmicutes and regulate gene expression during transcription. However, Gram-negative species such as Fusobacteria, Acidobacteria, and γ-proteobacteria also contain L box riboswitches that are predicted to regulate gene expression during translation initiation. Generally, L box riboswitches are dispersed around the chromosome upstream of lysine biosynthetic genes. The leader regions of large operons such as the *asd* gene cluster in *Thermatoga*
*maritima* also contain the L box riboswitch. In addition, several lysine transporters are also predicted to be under regulation by the L box riboswitch.

The initial L box secondary structure model (Fig. 1.12), derived from phylogenetic analysis, reveals conserved RNA structural motifs including an S-turn (or loop E) motif and an asymmetrical GA (or kink-turn) motif (Grundy *et al.*, 2003; Rodionov *et al.*, 2003; Sudarsan *et al.*, 2003). The *B. subtilis* *lysC* leader RNA utilizes alternative folding of RNA structures to modulate terminator formation based on cellular lysine concentrations. In the absence of lysine, an alternate antiterminator element forms and allows transcription to continue. In contrast, lysine biosynthetic genes in *E. coli* and other Gram-negative organisms contain a RBS-sequestering helix that promotes regulation at the level of translation initiation. With this mechanism of regulation, high lysine concentrations result in the formation of a helix that sequesters the RBS and prevents translation initiation. When lysine concentrations are low, the helix is not predicted to form, which makes the SD available for translation. *In vivo* and *in vitro* analysis confirmed that lysine directly promotes termination of the *lysC* leader RNA (Grundy *et al.*, 2003; Sudarsan *et al.*, 2003). The lysine analog aminoethylcysteine (AEC) required 10-fold higher concentrations to promote termination at a level similar to that promoted by lysine (Grundy *et al.*, 2003). Analysis of the binding affinity for lysine analogs demonstrated that the leader RNA is highly specific for lysine but exhibits a $K_D$ of $\sim 1 \mu M$. This is higher than other riboswitch RNAs that exhibit nM range affinities.
Figure 1.12. Sequence and predicted secondary structural model of the *B. subtilis* *lysC* leader RNA. Boxed letters and numbers indicate the terminator (T), antiterminator (AT), anti-antiterminator (AAT), and helices 1-6. Residues shown in red are found in 100% of the sequences analyzed and blue residues are found in greater than 80%. Green residues connected by dashed lines indicate sequences that show extensive covariation and form a tertiary interaction in the crystal structure (Garst et al., 2008; Serganov et al., 2008). Black text identifies conserved RNA motifs (Grundy et al., 2003). Adapted from Grundy et al. (2003).
Structural analysis of the *asd* leader from *T. maritima* uncovered the tertiary structure of the aptamer domain and identified elements important for lysine binding (Garst *et al*., 2008; Serganov *et al*., 2008). The crystal structure of the *asd* L box RNA reveals a complex structure distinct from other small-molecule binding riboswitches. The leader RNA adopts a compact structure with two coaxially stacked helical bundles radiating from the conserved five-way junction that forms an uneven “X” (Fig. 1.13). Helices 2 and 3 participate in a kissing interaction shown by mutational analysis to be essential to the regulatory function of the RNA (Blouin and Lafontaine, 2007). Helix 4 is anchored near the S-turn motif through interactions with conserved residues and the predicted kissing interaction was confirmed (Grundy *et al*., 2003; Garst *et al*., 2008; Serganov *et al*., 2008). The lysine molecule is bound within the conserved nucleotide core through an extensive network of nucleobase interactions (Garst *et al*., 2008; Serganov *et al*., 2008). In addition, a single K⁺ ion was buried in the lysine-binding pocket near the lysine carboxyl group (Fig. 1.14).
Figure 1.13. Global RNA structure of *T. maritima* L box aptamer domain. L box riboswitch structure in ribbon representation showing front and rotated by 60°. The bound lysine is in red and each RNA domain is numbered and labeled. The letters indicate paired regions (P) and loops (L) and labels are colored according to the corresponding RNA domain. Black text indicates conserved RNA motifs. Adapted from Serganov *et al.* (2009).
Further mutational analysis has highlighted the role of peripheral elements in lysine-dependent regulatory control. Briefly, the helix 2-loop 4 interaction is important for the organization of the lysine-binding pocket as deletion of the terminal loop resulted in deregulation (Blouin et al., 2011). It is predicted that formation of the kissing interaction between helices 2 and 3 is critical to the global RNA structure and is particularly required for the helix 2-loop 4 interaction. Finally, mutational analysis of the previously uncharacterized helix 6 indicated that this helix supports antitermination as disruption of the helix resulted in increased termination of the leader RNA.
1.2 Goals of this study

The goals of this study are focused on the structural and functional characterization of the L box riboswitch. At the onset of this project, few details were available on the mechanism of lysine recognition and binding, and the significance of the structure as it relates to the function of the riboswitch. Phylogenetic analysis by Grundy and co-workers (2003) offered a glimpse into the regulatory mechanism of the L box riboswitch. Their work demonstrated that the leader region is responsible for the premature termination of the B. subtilis lysC gene in response to cellular lysine concentrations. The initial goal of this project was to identify the sequence and structural features of the lysC leader RNA responsible for the lysine affinity. In our initial approach, we used an in vitro evolution method to identify lysC variants with the ability to terminate transcription in response to reduced lysine concentrations. This work identified specific sequences within the riboswitch RNA that are involved in lysine binding. Variants identified from the in vitro evolution were characterized based on the ability to modulate the RNA structure, bind lysine, and terminate transcription.

Previous work has indicated that the lysC leader RNA undergoes a conformational change in the presence of lysine. Modulation of the RNA structure is crucial to the function of riboswitches. This conformational change is predicted to affect expression of the downstream gene. The model of the L box leader RNA structural
transition, in response to lysine, identifies helix 1 as a key component to the structural change. Helix 1 is critical because it is one of few structural features found within the majority of known riboswitches. Analysis of the sequence of helix 1 in the lysC leader RNA revealed two base pairs that are highly conserved. The conservation at these positions prompted our analysis of the sequence of helix 1. Representative L box RNA variants, active at different levels of gene expression, were assayed to determine the concentration of lysine required for helix 1 formation. Mutations were introduced into the helix to examine the influence of the sequence of helix 1 on transcription termination. Although helix 1 is critical to lysine binding and formation of the terminator, we broadened our analysis to include additional structural features predicted to form in the presence of lysine. One such structural transition involves the docking of helix 4 onto helix 2. The long-range interaction between helices 2 and 4 is predicted to stabilize the binding pocket and global RNA structure. We therefore disrupted this interaction and challenged the leader RNA to terminate transcription in the presence of lysine. Together, these data offer a unique perspective and in-depth analysis of the sequence and structural requirements for the lysine-dependent structural transition of the L box riboswitch.

Finally, we used site-directed mutagenesis to characterize features of the B. subtilis lysC leader RNA responsible for lysine specificity and examine the structural elements of the lysine molecule required for recognition. The wild-type lysC leader RNA can recognize and discriminate between lysine and lysine analogs. Using mutational
analysis of the lysine-binding pocket, we identified leader RNA variants that exhibit changes in ligand recognition. These data demonstrate that \( \text{lysC} \) leader RNA specificity is the result of a series of distinct interactions between lysine and nucleotides that comprise the lysine-binding pocket, and provide insight into the molecular mechanisms employed by L box riboswitch RNAs to bind and recognize lysine.

In total, this work characterizes the sequence and structural features of the \( \text{lysC} \) leader RNA that are responsible for the modulation of gene expression. Analysis of the conserved residues in the structure emphasizes the importance of the structure-function relationship and highlights the significance of helix 1. Analysis of the long-range interactions presents an alternate view of the structural features required for termination, while mutation of the binding pocket reveals the molecular requirements for lysine recognition. In combination, these data offer a comprehensive view on the features of the lysine riboswitch that work in concert to regulate lysine biosynthetic genes in response to lysine.
CHAPTER 2

MUTATIONAL ANALYSIS OF L BOX STRUCTURAL ELEMENTS

2.1 Introduction

The L box leader RNA was the first amino acid-responsive riboswitch discovered. Since that time, two other riboswitches have been uncovered that respond to glycine and glutamine (Mandal et al., 2004b). The glycine and lysine RNAs represent two of the longest and most structurally complex riboswitches (Serganov et al., 2009). The aptamer domain of the L box leader RNA contains five helices coordinated around a highly conserved five-way junction. The glycine riboswitch contains two large aptamer domains that are adjacent to one another upstream of the regulatory structure. This is in contrast to the glutamine riboswitch that contains a relatively small aptamer domain with three base-paired regions and an average length of ~60 nt (Ames and Breaker, 2011).

The diverse aptamer domains among the amino acid binding riboswitches suggest that each leader RNA traverses a distinct folding pathway in response to the amino acid effector. This ligand-dependent conformational change is a requirement for regulation, as it is the conformational switching that determines the fate of the downstream gene. Currently, few details are available on the structural transitions of the lysC riboswitch that occur in response to lysine.
In vitro studies of the B. subtilis L box lysC leader revealed a series of structural rearrangements that are required for terminator formation. Biochemical analysis of the lysC leader RNA demonstrated that formation of the termination conformation also requires the stabilization of helix 1 (Grundy et al., 2003). RNase H cleavage confirmed that helix 1 formation only occurs in the presence of lysine. These results were instrumental to the development of a model of the lysine-dependent structural transition (Fig. 2.1). It was proposed that in the absence of lysine, the leader RNA forms an antiterminator structure that allows transcription to continue (Fig. 2.1). When lysine is present, however, the leader RNA adopts an alternate structure that sequesters the same sequences required to form the antiterminator through the formation of helix 1. Helix 1 formation stabilizes the anti-antiterminator and results in the formation of the transcriptional terminator (Fig. 2.1). Therefore, the presence of helix 1 is an indication of signal recognition and formation of the termination conformation. Despite the prevalence of helix 1 in a large number of the known riboswitches, few studies have focused on the role of helix 1 in regulatory function. It is clear, however, that helix 1 is crucial to the structural transition as it is often one of the main components in the ligand-dependent structural change. In some riboswitches, helix 1 is directly involved in the formation of the ligand-binding site, as seen in the B. subtilis yitJ S box riboswitch (Lu et al., 2010). Studies on the effects of helix 1 length in the B. subtilis lysC riboswitch suggest that the length of the helix is important for termination (Blouin et al., 2011). A 2 bp increase in the length of the helix resulted in disruption of lysine sensitivity and inhibited lysine-
dependent regulation. These results provide biochemical evidence that suggests that helix 1 may play a significant role in riboswitch regulation; however, little is known about the sequence determinants or structural features of the helix that are critical to function of the leader RNA. Helix 1 stability is crucial because it, in concert with ligand binding, determines the RNA conformation. We hypothesize that the stability of helix 1 is determined by the conserved sequences that dictate the propensity of the RNA to form the termination conformation. Therefore, we expect that mutation of the conserved residues in found helix 1 will alter the lysine-dependent response.

Figure 2.1. Lysine-dependent structural transition. Model of the *B. subtilis* lysC leader RNA lysine-dependent structural transition. In the absence of lysine, formation of the antiterminator (AT) allows readthrough. In the presence of lysine, the residues that form the 5’ side of the antiterminator (shown in green) are now sequestered in helix 1 of the antiterminator (AAT), releasing residues (shown in blue) that form the 5’ side of the terminator helix (T).
Recent X-ray crystallography studies have uncovered the global structure of the aptamer domain of the *T. maritima asd* leader RNA in the presence and absence of lysine (Garst *et al.*, 2008; Serganov *et al.*, 2008). This analysis indicates that the L box aptamer domain consists of three sets of co-linearly stacked helices stabilized by two long-range intramolecular interactions (Fig.1.13). Helices 1 and 2 are coaxially stacked and are parallel to coaxially stacked helices 4 and 5. The helical bundles form a structure similar to a disproportionate “X” that intersects at the lysine-binding pocket. This large structure is held together by helix 2 that participates in a kissing interaction with helix 3. This interaction helps to stabilize the global RNA structure. A second interaction with helix 2 occurs through the docking of helix 4 near the S-turn motif. Finally, helix 1 stabilizes the core of the molecule and helps to form the lysine-binding site. Although the leader RNA contains less sequence conservation as compared to other riboswitch RNAs, the majority of conserved nucleotides surround the lysine molecule in the binding pocket. A K⁺ ion was observed in the lysine-binding pocket wedged between the carboxyl group of the lysine molecule and the leader RNA (Serganov *et al.*, 2008). Crystallization of the bound and unbound L box aptamer domains resulted in two structures that were virtually identical, based on two independent analyses. Only minor structural differences were observed, which were localized to helix 1. Thus, both studies concluded that the leader RNA exists in a bound-like structure in the absence of lysine and undergoes only minor structural changes upon lysine binding. This is in contrast to the majority of known
riboswitches in which ligand binding results in significant structural modulation of the ligand-binding domain.

The crystal structure of the L box leader RNA has revealed two long-range tertiary interactions in the aptamer domain that involve helix 2. These interactions are predicted to be essential for lysine-dependent regulation. Initial phylogenetic analysis of the L box leader RNA revealed covariation between the terminal loops of helices 2 and 3. It was predicted that the two loops were involved in a tertiary interaction (Fig.1.12) (Grundy et al., 2003; Sudarsan et al., 2003). Later, crystal structure analysis revealed that the two loops participate in a kissing interaction that helps to stabilize the helices. The interaction between helices 2 and 3 is facilitated by the adjacent GA (kink-turn) motif. This motif was previously identified by phylogenetic analysis and is found in several other riboswitches such as the T box and S box leader RNAs. Although it does not fit the GA-motif consensus sequence, the structural element in the *T. maritima asd* L box riboswitch reverses the orientation of the loop which imparts a 120° turn in the backbone of helix 2 and allows presentation of the nucleotides to helix 3 for the kissing interaction (Garst et al., 2008; Serganov et al., 2008). Transcription of *B. subtilis lysC* leader RNA variants with mutations in loops 2 and 3 or disruption of the GA-motif demonstrated that formation of the loop-loop kissing interaction is required for efficient lysine-dependent termination (Blouin and Lafontaine, 2007).
The helix 2-loop 4 interaction is predicted to stabilize the lysine binding pocket. The tertiary structure illustrates that two highly conserved A residues in the terminal loop of helix 4 interact with two invariant G residues adjacent to the S-turn motif (Garst et al., 2008; Serganov et al., 2008). Detailed molecular analysis of the interaction indicates that the terminal loop of helix 4 employs an unusual helix docking mechanism (Serganov et al., 2008). The Watson-Crick face of each nucleotide in the terminal loop of helix 4 directly interacts with the minor groove of helix 2. This is in contrast to helical docking in other large RNAs that employ the more common A-minor motif for docking interactions (Nissen et al., 2001; Garst et al., 2008; Serganov et al., 2008). Unlike the kissing interaction between helices 2 and 3, biochemical data suggest that the helix 2-loop 4 interaction occurs upon lysine binding and is not present in the lysine-free structure (Serganov et al., 2008). Changes to the length of helix 4 and deletion of the terminal loop resulted in a loss of lysine-dependent termination (Blouin et al., 2011).

To assess the role of the tertiary intramolecular interactions of the *B. subtilis* *lysC* leader RNA, we targeted regions predicted to be directly involved in the formation of the lysine-bound structure. Oligonucleotide-directed RNase H cleavage was used to determine the concentration of lysine required for helix 1 stabilization. Two RNA representatives were analyzed, the *B. subtilis lysC* leader RNA, which regulates gene expression at the transcriptional level, and the *E. coli lysC* leader RNA, which is predicted to regulate during translation initiation. Additionally, conserved residues in
helix 1 were mutagenized to assess the effects of the sequence on lysine-dependent termination. Results of the mutational analysis uncovered sequences essential to the formation of helix 1. Helix 1 variants were also tested in vivo to assess the physiological effects of the mutations. Finally, variant RNAs with mutations in the terminal loop of helix 4 were tested to determine the effects on lysine-dependent termination when the helix 2-loop 4 interaction was disrupted.

2.2 Materials and Methods

2.2.1 Generation of DNA templates

Templates for in vitro transcription by B. subtilis RNA polymerase (RNAP) were generated by fusion of the B. subtilis glyQS promoter sequence to the lysC leader RNA sequence. PCR was carried out using DNA oligonucleotide primers that included the glyQS promoter and hybridized within the lysC leader region. A transcriptional fusion plasmid, pFG328, with a wild-type lysC insertion (Grundy et al., 2003) was the DNA source for PCR amplification using Taq DNA polymerase per the manufacturer’s instructions (Invitrogen, Carlsbad, CA). The resulting PCR fragment contained the lysC leader template downstream of the glyQS promoter such that transcription would start at +17 relative to the lysC transcription start-site, and included the lysC sequence extending 14 nt downstream of the termination site. The transcription start-site was designed to allow transcription initiation with the dinucleotide ApC and a halt at G42 (relative to the native lysC transcript) during transcription in the absence of
CTP. The 3’ end of the construct included an additional 106 bp of random sequence to allow resolution between terminated (252 nt) and readthrough (372 nt) products. PCR products were purified using a QIAquick PCR clean up kit (Qiagen, Chatsworth, CA) and sequenced by Genewiz (South Plainfield, NJ).

DNA templates used for phage T7 RNAP transcription of the *B. subtilis* *lysC* leader RNA were generated using complementary pairs of overlapping DNA oligonucleotides (Table 2.1), as previously described (McDaniel *et al.*, 2005; Yousef *et al.*, 2005). Briefly, the 5’ pair contained the phage T7 RNAP promoter sequence fused to the +18 position of the *lysC* leader region. The remaining complementary pairs contained the wild-type *lysC* sequence each with a 5-nt 3’ overhang complementary to the 5’ region of the adjacent pair. The terminal pair contained the 3’ leader RNA sequence that ended at +234, 3 nt prior to the 5’ position of the transcription terminator. This position was selected so that helix 1 formation could be monitored without the competing antiterminator structure. Each internal oligonucleotide pair was phosphorylated using T4 polynucleotide kinase (New England Biolabs, Beverly, MA). The pairs were then mixed, incubated at 95°C, and slow cooled to room temperature for annealing. The paired oligonucleotides were ligated using T4 DNA ligase per the manufacturer’s instructions (New England Biolabs, Beverly, MA). The resulting DNA template was amplified using the flanking 5’ and 3’ DNA oligonucleotides as primers for PCR using *Taq* DNA polymerase (Invitrogen, Carlsbad, CA). The final template was purified and sequenced as described above.
Table 2.1. DNA oligonucleotide primers for the *B. subtilis* lysC leader

<table>
<thead>
<tr>
<th>Primer</th>
<th>Position</th>
<th>Sequence (5' to 3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>LysC T7 US</td>
<td>T7 prom-+18</td>
<td>TAATACGACTCACTATAGGTATATGGTGGAAGTAGTGAGGTCGCG</td>
</tr>
<tr>
<td>LysC CассA wt</td>
<td>+18-59</td>
<td>GGTATTTATATGGTGGAAGTAGTGAGGTCGCGAATCTCAAGATAGT</td>
</tr>
<tr>
<td>LysC CассA wt rc</td>
<td>+18-64</td>
<td>TCTTGAAGTTTCGACCTATCTTTCACCATAAAAC</td>
</tr>
<tr>
<td>LysC CассB wt</td>
<td>55-104</td>
<td>CCTTTTGCAAGCTTTTCACAGATTCATCCCTTTCACAAAGGCA</td>
</tr>
<tr>
<td>LysC CассD wt</td>
<td>105-144</td>
<td>GTGCGCGAGCAAGCTAAATACCGTTGTATTTTGTG</td>
</tr>
<tr>
<td>LysC CассD wt rc</td>
<td>100-139</td>
<td>AATAATACGATGGGGTTTTATTTTGTGCGAGCGCTTCC</td>
</tr>
<tr>
<td>LysC CассE</td>
<td>145-178</td>
<td>GCGGCTGATGAAATATGTAGGGCTGCAAAGAA</td>
</tr>
<tr>
<td>LysC CассF wt</td>
<td>179-212</td>
<td>ATACCTTTCTGGAGGGGCTACTCTGGTGTTCCATA</td>
</tr>
<tr>
<td>LysC CассF wt rc</td>
<td>174-207</td>
<td>ACAACGAGATAGCCCTCCAAGAAAATGTATTG</td>
</tr>
<tr>
<td>LysC CассG</td>
<td>213-234</td>
<td>TCAATTAAGCTTTCAATGATTAGATATGGA</td>
</tr>
<tr>
<td>LysC CассG rc</td>
<td>208-229</td>
<td>TAATATCATAAAATGTATTAGGCA</td>
</tr>
<tr>
<td>LysC CассBsp3' helix1</td>
<td>201-234</td>
<td>TCAATTAAGCTTTCAATGATTAGATATGGA</td>
</tr>
<tr>
<td>LysC RNH2</td>
<td>193-204</td>
<td>ACGAGATAGCC</td>
</tr>
<tr>
<td>LysC glyQS U35A</td>
<td>glyQS prom-+56</td>
<td>TGGGTTATATGGTGGAAGTAGGGGCTGCAATCTCAAGAGATAGT</td>
</tr>
<tr>
<td>LysC glyQS U35A RC</td>
<td>glyQS prom-+56</td>
<td>ACTCTTTGAAGTTTCGACACCTTCTTTCACCATATAACACTCA</td>
</tr>
<tr>
<td>LysC glyQS U35C</td>
<td>glyQS prom-+56</td>
<td>GACAGCCTTACATTTACTGCAATGCTGCAAAGGCA</td>
</tr>
<tr>
<td>LysC glyQS U35C RC</td>
<td>glyQS prom-+56</td>
<td>ACTCTTTGAAGTTTCGACACCTTCTTTCACCATATAACACTCA</td>
</tr>
<tr>
<td>LysC glyQS U35G</td>
<td>glyQS prom-+56</td>
<td>GACAGCCTTACATTTACTGCAATGCTGCAAAGGCA</td>
</tr>
<tr>
<td>LysC glyQS U35G RC</td>
<td>glyQS prom-+56</td>
<td>ACTCTTTGAAGTTTCGACACCTTCTTTCACCATATAACACTCA</td>
</tr>
<tr>
<td>LysC glyQS U355 GC</td>
<td>glyQS prom-+56</td>
<td>GACAGCCTTACATTTACTGCAATGCTGCAAAGGCA</td>
</tr>
<tr>
<td>LysC glyQS U355 GC RC</td>
<td>glyQS prom-+56</td>
<td>ACTCTTTGAAGTTTCGACACCTTCTTTCACCATATAACACTCA</td>
</tr>
<tr>
<td>LysC glyQS A36C</td>
<td>glyQS prom-+56</td>
<td>GACAGCCTTACATTTACTGCAATGCTGCAAAGGCA</td>
</tr>
<tr>
<td>LysC glyQS A36C RC</td>
<td>glyQS prom-+56</td>
<td>ACTCTTTGAAGTTTCGACACCTTCTTTCACCATATAACACTCA</td>
</tr>
<tr>
<td>LysC glyQS A36G</td>
<td>glyQS prom-+56</td>
<td>GACAGCCTTACATTTACTGCAATGCTGCAAAGGCA</td>
</tr>
<tr>
<td>LysC glyQS A36G RC</td>
<td>glyQS prom-+56</td>
<td>ACTCTTTGAAGTTTCGACACCTTCTTTCACCATATAACACTCA</td>
</tr>
<tr>
<td>LysC glyQS A36U</td>
<td>glyQS prom-+56</td>
<td>GACAGCCTTACATTTACTGCAATGCTGCAAAGGCA</td>
</tr>
<tr>
<td>LysC glyQS A36U RC</td>
<td>glyQS prom-+56</td>
<td>ACTCTTTGAAGTTTCGACACCTTCTTTCACCATATAACACTCA</td>
</tr>
<tr>
<td>LysC glyQS U197A</td>
<td>176-217</td>
<td>GAAATCATTCTTCTGGAGGGCTCTCTCGTTGTATTAATCAT</td>
</tr>
<tr>
<td>LysC glyQS A198C</td>
<td>176-217</td>
<td>GAAATCATTCTTCTGGAGGGCTCTCTCGTTGTATTAATCAT</td>
</tr>
<tr>
<td>LysC glyQS A198C RC</td>
<td>176-217</td>
<td>GAAATCATTCTTCTGGAGGGCTCTCTCGTTGTATTAATCAT</td>
</tr>
<tr>
<td>LysC glyQS A198U</td>
<td>176-217</td>
<td>GAAATCATTCTTCTGGAGGGCTCTCTCGTTGTATTAATCAT</td>
</tr>
<tr>
<td>LysC glyQS A198U RC</td>
<td>176-217</td>
<td>GAAATCATTCTTCTGGAGGGCTCTCTCGTTGTATTAATCAT</td>
</tr>
<tr>
<td>LysC glyQS A36C</td>
<td>glyQS prom-+56</td>
<td>GACAGCCTTACATTTACTGCAATGCTGCAAAGGCA</td>
</tr>
<tr>
<td>LysC glyQS A36C RC</td>
<td>glyQS prom-+56</td>
<td>ACTCTTTGAAGTTTCGACACCTTCTTTCACCATATAACACTCA</td>
</tr>
<tr>
<td>LysC glyQS A36G</td>
<td>glyQS prom-+56</td>
<td>GACAGCCTTACATTTACTGCAATGCTGCAAAGGCA</td>
</tr>
<tr>
<td>LysC glyQS A36G RC</td>
<td>glyQS prom-+56</td>
<td>ACTCTTTGAAGTTTCGACACCTTCTTTCACCATATAACACTCA</td>
</tr>
<tr>
<td>LysC glyQS A36U</td>
<td>glyQS prom-+56</td>
<td>GACAGCCTTACATTTACTGCAATGCTGCAAAGGCA</td>
</tr>
<tr>
<td>LysC glyQS A36U RC</td>
<td>glyQS prom-+56</td>
<td>ACTCTTTGAAGTTTCGACACCTTCTTTCACCATATAACACTCA</td>
</tr>
</tbody>
</table>

(continued)
Table 2.1 (continued)

<table>
<thead>
<tr>
<th>Primer</th>
<th>Position</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LysC glyQS U197G RC</td>
<td>176-217</td>
<td>ATGATATGAAACAGAGATCGCCCTCAAGAAAATGATTTC</td>
</tr>
<tr>
<td>LysC A156C</td>
<td>132-175</td>
<td>GTATATTTTGTGCGTGCTGATTTGCAAAATGTAAGGCTGCA</td>
</tr>
<tr>
<td>LysC A156C RC</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A156G</td>
<td>132-175</td>
<td>GTATATTTTGTGCGTGCTGATTTGCAAAATGTAAGGCTGCA</td>
</tr>
<tr>
<td>LysC A156G RC</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A16T</td>
<td>132-175</td>
<td>GTATATTTTGTGCGTGCTGATTTGCAAAATGTAAGGCTGCA</td>
</tr>
<tr>
<td>LysC A16T RC</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A17C</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A17C RC</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A17G</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A17G RC</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A17T</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A17T RC</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A18C</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A18C RC</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A18G</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A18G RC</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A18T</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
<tr>
<td>LysC A18T RC</td>
<td>132-175</td>
<td>TTGACAGCCCTACATTATGCAATGCACCAGCAGCAAAATAC</td>
</tr>
</tbody>
</table>

DNA templates for T7 RNAP transcription of the *E. coli* *lysC* leader were generated by PCR with *E. coli* PALΔSΔUTR chromosomal DNA as the template, using a 5’ primer (T7prom+18; Table 2.2) that contained the T7 RNAP promoter sequence fused to position +18 of the *E. coli* *lysC* leader region. Position +232 was selected as the endpoint of the *lysC* fragment and 20 random nucleotides were added downstream to allow resolution of cleaved and full-length transcripts. The resulting fragment was purified and sequenced as described above.

Table 2.2. DNA oligonucleotide primers for the *E. coli* *lysC* leader

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>T7 prom+18</td>
<td>TAATACGACTCACTATAGGGGAGCTGAGAAGAG</td>
</tr>
<tr>
<td>E.coli hlx1short</td>
<td>CGATATGCACGCTAGCCAGAAGGTACCTGCAGC</td>
</tr>
<tr>
<td>RNH13</td>
<td>CCAGAAG</td>
</tr>
</tbody>
</table>
2.2.2 Site-directed mutagenesis

The transcriptional fusion vector pFG328 containing the wild-type lysC DNA transcript (Grundy et al., 2003), as described above, was used as the template for oligonucleotide-directed mutagenesis. DNA oligonucleotides containing the desired mutations (Table 2.1) were designed as primers for PCR amplification of pFG328-lysC DNA using *Pfu* DNA polymerase (Stratagene, La Jolla, CA). The resulting products were subjected to digestion with Dpn I (New England Biolabs, Beverly, MA) to remove the starting wild-type template. The resulting template was introduced into *E. coli* XL-2 blue ultracompetent cells by transformation per the manufacturer’s instructions (Stratagene, La Jolla, CA). The plasmid DNA was isolated using a Wizard prep kit (Promega, Madison, WI) and sequenced to confirm the mutations (Genewiz, South Plainfield, NJ). The final constructs contained the glyQS promoter sequence upstream of the lysC leader RNA with the desired mutations, and were used as the DNA template for *in vitro* transcription.

2.2.3 RNase H cleavage assay

RNase H cleavage was carried out as described by Grundy *et al.* (2003). The DNA templates for T7 RNAP transcription of the *B. subtilis* or *E. coli* lysC leader regions were generated as described above. Transcription was carried out using a MEGAshortscript T7 RNAP transcription kit (Ambion, Austin, TX) in the presence of 0.5 µM [α-32P]-UTP (GE Healthcare; 800 Ci/mmol [30 TBq/mmol]) for radiolabelling.
RNAs were transcribed in the presence or absence of varying concentrations of lysine or AEC, as indicated, for 30 min at 37°C. After transcription, a DNA oligonucleotide (47 µM) complementary to positions 193-204 of the *B. subtilis lysC* leader RNA, or 202-214 of the *E. coli lysC* RNA, was added and incubated for 5 min at 37°C. RNase H (0.45 U; Ambion, Austin, TX) was then added to the reaction and incubated 10 min at 37°C. The reactions were stopped by phenol-chloroform extraction. The resulting RNA products were resolved by denaturing PAGE and visualized by PhosphorImager analysis (Molecular Dynamics). The percentage of RNAs protected from cleavage was calculated as the amount of full length RNAs relative to the total amount of RNA transcribed in each reaction. The reactions were performed in duplicate, and reproducibility was ±10%.

### 2.2.4 *In vitro* transcription

Single round transcription of the wild-type *lysC* leader RNA was performed as previously described (Grundy *et al.*, 2003). A reaction mixture of 20 mM Tris-HCl, pH 8, 20 mM NaCl, 10 mM MgCl₂, 100 µM EDTA, 150 µM ApC (Sigma), 2.5 µM GTP and ATP, 0.75 µM UTP (0.25 µM, 800 Ci/mmol), [α-³²P]-UTP, DNA template (10 nM), and His-tagged purified *B. subtilis* RNAP (6 nM) (Qi and Hulett, 1998) was incubated for 15 min at 37°C. CTP was excluded from the reaction mixture to generate a transcriptional halt at +42 (relative to the native *lysC* transcription start-site). After incubation, heparin was added to block reinitiation and synchronize transcript synthesis. Lysine was added to a final concentration of 1.5 mM and elongation was resumed by the addition of 10 µM
rNTPs and 40 mM MgCl₂ followed by incubation for 15 min at 37°C. The transcription reactions were stopped by phenol/chloroform extraction and the products were resolved by denaturing PAGE and visualized by PhosphorImager analysis (Molecular Dynamics). Percent termination represents terminated transcripts relative to the total amount of RNA transcribed; this value was plotted as a function of ligand concentration. Non-linear regression analysis was used to determine the concentration required for half-maximal termination, Trm½ (Prism Graph, GraphPad Software). Reactions were performed in duplicate, and reproducibility was ±5%.

2.2.5 lysC-lacZ transcriptional fusions and β-galactosidase assays

The *B. subtilis* BR151 (*lys-3 metB10 trpC2*) strain was used to examine the in vivo regulatory effects of the *lysC* leader RNA on reporter gene expression. A single copy of the transcriptional fusion vector pFG328, containing the *B. subtilis lysC* leader sequence fused to *lacZ*, was integrated into the BR151 chromosome as described by Grundy *et al.* (2003). Cells harboring the appropriate fusion construct were selected and grown to mid-exponential phase in Spizizen minimal medium (Anagnostopoulous and Spizizen, 1961). The cells were then harvested, separated into two cultures and resuspended in minimal media. For cells grown in low lysine conditions, the media was supplemented with 50 µg/mL of methionine and tryptophan. To monitor expression when lysine concentrations are high, the media was supplemented with 50 µg/mL of methionine and tryptophan, and lysine was added the culture to a final concentration of
100 µg/mL. The cells were grown at 37°C for an additional four hours and samples were collected each hour to measure β-galactosidase activity as described by Miller (1972). Experiments were performed in duplicate and reproducibility was ±10%.

2.3 Results

2.3.1 Lysine promotes stabilization of helix 1

The current L box model predicts that formation of helix 1 stabilizes the anti-antiterminator in the presence of lysine. Oligonucleotide-directed RNase H cleavage was used to monitor the formation of helix 1 in the presence of lysine or AEC. RNase H is an endonuclease that specifically cleaves RNA-DNA hybrids. T7 RNAP was used to synthesize a radiolabelled *B. subtilis* lysC leader RNA in the presence or absence of lysine. A DNA oligonucleotide complementary to the 3’ side of helix 1 was added followed by incubation with RNase H. Transcripts that formed helix 1 were protected from cleavage, whereas transcripts unable to form helix 1 were cleaved by RNase H. In the absence of lysine, the wild-type *lysC* leader RNA was cleaved by RNase H (Fig. 2.2A). This is consistent with the model that the unliganded structure does not form helix 1. In the presence of lysine, helix 1 is protected from RNase H cleavage, which indicates that formation of helix 1 is a direct consequence of lysine binding.
Figure 2.2. Helix 1 cleavage of the *B. subtilis lysC* leader RNA by RNase H. A) Denaturing PAGE of RNase H cleavage products. A DNA oligonucleotide (*lysC* RNH2) complementary to the 3’ region (positions 193-204) of the *B. subtilis lysC* leader RNA was added to radiolabelled RNA in the presence of decreasing lysine or AEC concentrations. RNase H was added for the cleavage of DNA-RNA hybrids. Cleavage products were visualized by denaturing PAGE followed by PhosphorImager analysis. B) Plot of helix 1 cleavage protection as a function of ligand concentration. Non-linear regression analysis was used to determine the concentration required for half-maximal protection, $P_{1/2}$. FL, full length transcripts; C transcripts cleaved by RNase H. $P_{1/2}$, half-maximal protection. The reactions were performed in duplicate, and reproducibility was ±10%.

Lysine titration in the RNase H assay demonstrated that helix 1 formation was dose-dependent. As the concentration of lysine was increased, helix 1 protection increased which is consistent with previous RNase H analysis of the *lysC* leader RNA (Fig. 2.2A) (Grundy et al., 2003). The RNase H assay was also performed in the presence of AEC. AEC is a lysine analog that contains a sulfur substitution in the carbon side chain. The *B. subtilis lysC* leader RNA required higher concentrations of AEC than lysine for helix 1 protection. Non-linear regression analysis of leader RNAs transcribed
in the presence of lysine required 230 µM for half-maximal protection (Fig. 2.2B). In the presence of AEC, there was a 2-fold increase in the concentration required for half-maximal protection, as compared to lysine. These findings are consistent with previous \textit{in vitro} transcription results that demonstrated that the leader RNA is specific for lysine and has a decreased sensitivity to AEC (Grundy et al., 2003). Analysis of helix 1 formation of the \textit{E. coli} \textit{lysC} leader RNA was similar to the \textit{B. subtilis} \textit{lysC} leader, and exhibited an increase in helix 1 protection as the concentration of lysine or AEC was increased (Fig. 2.3A). The concentrations required for half-maximal protection of the \textit{E. coli} leader RNA were 360 µM for lysine and 830 µM for AEC (Fig. 2.3B).
Figure 2.3. Helix 1 cleavage of the *E. coli* lysC leader RNA by RNase H. (A) Denaturing PAGE of RNase H cleavage products. A DNA oligonucleotide (*lysC* RNH2) complementary to the 3’ region (positions 202-214) of the *E. coli* lysC leader RNA was added to radiolabelled RNA in the presence of decreasing lysine or AEC concentrations. RNase H was added for the cleavage of DNA-RNA hybrids and products were visualized by denaturing PAGE followed by PhosphorImager analysis. (B) Plot of helix 1 cleavage protection as a function of ligand concentration. Non-linear regression analysis was used to determine the concentration required for half-maximal protection, $P_{1/2}$. FL, full length transcripts; C transcripts cleaved by RNase H. $P_{1/2}$, half-maximal protection. The reactions were performed in duplicate, and reproducibility was ±10%.

The *E. coli* lysC results were similar to those for the *B. subtilis* leader RNA, where a 2-fold decrease in protection of helix 1 was observed in the presence of AEC as compared to lysine. The lysine dependence of helix 1 formation in a transcriptionally
active riboswitch, as compared to an L box riboswitch that regulates during translation, indicated that lysC leader RNAs require similar ligand concentrations for structural modulation. However, the lysC leader from B. subtilis exhibited increased protection of the leader RNA at lower ligand concentrations as compared to protection exhibited by the E. coli leader RNA. This may be an indication of a differential ligand affinity between the two RNAs in which the B. subtilis leader RNA can form the termination conformation at lower ligand concentrations as compared to the E. coli L box riboswitch.

2.3.2 Mutational analysis of helix 1

A recent study of helix 1 suggested that increased stability of helix 1 has a direct effect on termination (Blouin et al., 2010). lysC leader RNA variants with helix 1 extensions were analyzed in vitro. An increase in the length of the helix (4 bp) caused an increase in termination in both the presence and absence of lysine, which suggests that helix 1 is involved in the stability of the two RNA conformations. In this study, we have focused on the sequence of helix 1. Phylogenetic analysis of L box leader RNAs reveals variation in both the length and base pair composition of helix although two base pairs near the base of the helix are highly conserved (Barrick and Breaker, 2007).

In vitro transcription of the wild-type lysC leader RNA exhibited a 3-fold increase in termination in the presence of lysine (1.5 mM) as compared to termination in the absence of lysine (Fig. 2.4). Variants of the lysC leader RNA harboring mutations at the U35-A198 position were constructed and each variant was transcribed and challenged to
terminate transcription in the presence of 1.5 mM lysine. The A35-U198 exhibited a 3-fold increase in termination when transcribed in the presence as compared to the absence of lysine. For the A35-U198 variant, it is interesting that reversing the base pair did not have a significant effect on termination. This is in contrast to the opposing G-C/C-G pairs, which had different effects on termination. The G35-C198 variant exhibited a 3-fold increase in termination, which is comparable to the increase in termination of the wild-type construct in the presence of lysine. Termination of the C35-G198 variant in the absence of lysine increased 3-fold as compared to wild type. In the presence of lysine, the C-G variant exhibited increased termination relative to wild type (91% and 76%, respectively). These results suggest that a C-G base pair has a direct effect on the ligand-independent structure of the leader RNA.

Substitution of the wild-type sequence with G35-U198 or U35-G198 resulted in a reduction in termination, as compared to wild type (Fig. 2.4A; lanes 9-12). Although both variants exhibited increased termination, the U-G mutation resulted in a 5-fold increase in termination in the presence as compared to the absence of lysine, whereas the G-U variant exhibited a 3-fold increase in termination when lysine was present. Mutations that disrupted base pairing at U35-A198 resulted in a significant decrease in termination in the presence and absence of lysine (C35-A198, A35-C198 and G35-A198, Fig. 2.4A; lanes 13-18). Mutational analysis of the adjacent A36-U197 base pair was also performed. Phylogenetic analysis of the leader RNA suggests that this base pair is
more conserved than U35-A198 (Barrick and Breaker, 2007). In the crystal structure, A36-U197 is positioned at the base of the lysine-binding pocket.

Figure 2.4. *In vitro* transcription of U35-A198 variant *lysC* leader RNAs. (A) *In vitro* transcription of leader RNAs with mutations in U35-A198 in the presence of 1.5 mM lysine. **FL**, full-length transcripts. **T**, terminated transcripts. (B) Bar graph of percent termination for each construct. Open bars represent transcription termination in the absence of lysine and filled bars represent transcription in the presence of 1.5 mM lysine. Percent termination is the fraction of RNAs terminated relative to the total amount of RNA in each transcription reaction. Reactions were performed in duplicate, and reproducibility was ±5%.
The proximity to the binding pocket suggests that mutation of A36-U197 might be more detrimental to the regulatory function of the leader RNA than U35-A198. Substitution of the wild-type A36-U197 base pair with U36-A197 caused a reduction in termination. Unlike the A35-U198 variant in which reversal of the base pair was not tolerated, the U36-A197 variant was able to promote lysine-dependent termination albeit with reduced efficiency as compared to wild type. The G36-C197 variant showed lysine-dependent termination similar to that of wild type, with a 2.5-fold increase in termination in the presence of lysine, as compared to termination in the absence of lysine. In contrast, the C36-G198 mutations caused an increase in lysine-independent termination. In the absence of lysine, termination was 2.5-fold higher than wild-type termination. When the variant was transcribed in the presence of lysine, an 18% increase in termination was observed as compared to wild type. The lysine-independent increase in termination suggests that the C-G base pair results in increased stability of helix 1, and causes the RNA to favor the terminated structure in the absence of lysine. Termination of the G36-U197 variant was similar to that of G35-U198 and demonstrated a reduction in termination efficiency as compared to wild type. The variant RNA exhibited an increased lysine response, which resulted in a 9-fold increase in termination when transcribed in the presence of lysine. This is 3-fold higher than the wild type increase in termination when transcribed in the presence vs. the absence of lysine. This increase in termination in the presence of lysine suggests that the G36-U197 mutation causes increased lysine sensitivity; however, the reduced termination efficiency may indicate
that the mutations destabilize the structure of helix 1. In total, the *in vitro* termination results support our hypothesis; mutations at the A36-U197 position are more detrimental than mutations at U35-A198 as there was only one variant that resulted in termination similar to that of wild type.
Figure 2.5. *In vitro* transcription of A36-U197 variant *lysC* leader RNAs. (A) PAGE of single round transcription reactions of leader RNAs with mutations in A36-U197 in the presence of 1.5 mM lysine. FL, full-length transcripts. T, terminated transcripts. (B) Bar graph of percent termination for each construct. Open bars represent transcription termination in the absence of lysine and filled bars represent transcription in the presence of 1.5 mM lysine for each variant. Percent termination is the fraction of RNAs terminated relative to the total amount of RNA for each transcription reaction. Reactions were performed in duplicate, and reproducibility was ±5%.

These results also confirmed that base pairing at A36-U197 is required for efficient lysine-dependent termination. Disruption of the base pair at the 36-197 position
caused an increase in antitermination. The C36-A197 variant, which is predicted to disrupt the base pairing, exhibited 45% termination in the presence and absence of lysine (Fig. 2.5B). This suggests that the identity of the base pair at the 36-197 position is important to the stability of the lysine-independent structure.

2.3.3 *In vivo expression of variant lysC-lacZ transcriptional fusions*

To examine the effects of the helix 1 mutations on gene expression, additional *in vivo* analysis was performed. *In vitro* transcription revealed that alteration of the conserved sequence in helix 1 has a direct effect on regulation. However, it was not clear if the mutations would have a similar effect *in vivo*. Thus, the *B. subtilis* lysC leader RNA was positioned downstream of the *B. subtilis* glyQS promoter and fused to a promoter-less lacZ gene. A single copy of this transcriptional fusion was inserted into the chromosome of the *B. subtilis* lysine auxotrophic strain BR151. β-galactosidase activity was monitored when cells were grown in the absence or presence of 100 µg/mL of lysine. Previous analyses of L box leader RNA transcriptional fusions have shown repression of gene expression when lysine levels are high and induction of expression under low lysine conditions (Grundy *et al.*, 2003; Sudarsan *et al.*, 2003). Expression of the wild-type lysC-lacZ fusion was repressed 7-fold when cells were grown in the presence of 100 µg/mL of lysine (Table 2.3), whereas expression was induced in the absence of lysine.
Table 2.3. Expression of lysC-lacZ fusions in B. subtilis.

<table>
<thead>
<tr>
<th>Fusion b</th>
<th>-Lys</th>
<th>+ Lys</th>
<th>Induction ratio c</th>
</tr>
</thead>
<tbody>
<tr>
<td>WT</td>
<td>89 ± 6.6</td>
<td>12 ± 0.54</td>
<td>7.4</td>
</tr>
<tr>
<td>U35-A198</td>
<td>45 ± 14</td>
<td>9.4 ± 1.2</td>
<td>4.8</td>
</tr>
<tr>
<td>G35-C198</td>
<td>16 ± 0.35</td>
<td>6.9 ± 0.42</td>
<td>2.3</td>
</tr>
<tr>
<td>C35-G198</td>
<td>10 ± 3.2</td>
<td>5.9 ± 1.1</td>
<td>1.7</td>
</tr>
<tr>
<td>C35-A198</td>
<td>170 ± 7.2</td>
<td>250 ± 12</td>
<td>0.68</td>
</tr>
<tr>
<td>A36-U197</td>
<td>98 ± 6.5</td>
<td>30 ± 4.9</td>
<td>3.2</td>
</tr>
<tr>
<td>G36-C197</td>
<td>40 ± 3.8</td>
<td>10 ± 2.2</td>
<td>4.0</td>
</tr>
<tr>
<td>C36-G197</td>
<td>21 ± 3.7</td>
<td>10 ± 0.9</td>
<td>2.1</td>
</tr>
</tbody>
</table>

a Cells were grown in minimal medium (Anagnostopoulos and Spizizen, 1961), harvested, split and grown in the absence (-Lys) or presence of 100 μg/mL of lysine(+Lys). β-galactosidase activity is shown in Miller units (Miller, 1972) for samples taken 3 hours after harvesting. b Transcriptional fusions consisted of either wild-type (WT) or the indicated mutant positioned downstream of the B. subtilis glyQS promoter fused to a promoterless lacZ gene from E. coli. c Induction ratio of β-galactosidase activity in the absence of lysine divided by the activity in the presence of lysine. Experiments were performed in duplicate and reproducibility was ±10%

The wild-type lysC leader RNA expression profile was consistent with previously published results (Grundy et al., 2003; Sudarsan et al., 2003). The highly conserved residues at U35-A198 and A36-U197 were targeted for mutagenesis. Expression of the A35-U198 variant was reduced 2-fold in the absence of lysine, as compared to wild type. The reduction in expression might be due to disruption of additional interactions that may occur in the absence of helix 1 formation. The G35-C198 variant also exhibited lysine-
dependent repression; however, expression was 5.5-fold lower than wild type in the absence of lysine and 2-fold lower when cells were grown in 100 μg/mL lysine. It is possible that the G35-C198 mutations offer increased stability to helix 1 with the substitution of a G-C pair for A-U. In addition, the 3’ side of helix 1 is also involved in the formation of the AT structure in the absence of lysine (Fig. 2.1). In the AT, a single nucleotide substitution of A198 with C results in an A-U to C-U swap. This base pair disruption may cause destabilization of the AT and subsequently support formation of the termination conformation. Expression of the C35-G198 variant in the absence of lysine exhibited a 9-fold reduction in basal gene expression; this reduction may also be the result of increased helix 1 stability through the substitution of a G-C pair for A-U. In contrast to the G35-C198 variant, the C35-G198 variant is not predicted to destabilize additional structures as the mutation results in a G-U substitution in the AT. Although it is not likely that a G-U pair disrupts AT formation, the base pair swap in the AT in conjunction with increased helix 1 stability may account for the reduction in the basal level of gene expression. U35-A198 mutations that maintained the base pairing in helix 1 caused a reduction in overall expression as compared to variants that disrupted base pairing. Disruption of the U35-A198 base pair caused constitutive expression of lacZ. The expression of the variants was at least 2-fold (in the absence of lysine) and 20-fold (in 100 μg/mL lysine) higher than wild-type expression under the same conditions (Table 2.3). This result is consistent with the previous in vitro transcription assays and suggests
that the formation of helix 1 is crucial to the regulatory function of the leader RNA *in vivo* and *in vitro*.

Expression of the A36-U197 variant was similar to wild type when cells were grown in the absence of lysine. When cells were grown in the presence of 100 μg/mL lysine, gene expression was repressed 3-fold. This is a 2.4-fold decrease in repression as compared to wild-type repression when cells were grown under the same conditions. When the U-A pair was substituted with either G-C or C-G, a 2- to 4-fold reduction in expression was observed in the absence of lysine. For the G36-C197 variant, expression was repressed 4-fold when lysine was present. Expression of the C36-G197 variant was repressed 2-fold, which is consistent with the *in vitro* transcription assays. The low basal gene expression observed *in vivo* for the G-C/C-G variants suggests that the substitution of the A36-U197 pair causes increased stability of the termination conformation. This could be the result of an increased helix 1 stability with the introduction of the more stable G-C/C-G pair. It is possible that the mutation allows the anti-antiterminator to more effectively compete with the antiterminator for structure formation.

**2.3.4 Mutational analysis of helix 4 terminal loop**

To determine the functional role of the helix 2-loop 4 interaction, we mutated the highly conserved A156 and A157 positions in the *lysC* leader RNA. Mutation of the conserved residues is predicted to disrupt the long-range tertiary interaction. We challenged the A156 and A157 variant RNAs to terminate transcription in the presence of
lysine. *In vitro* termination of the A156 variants indicated that substitution of position 156 caused a reduction in termination efficiency as compared to wild type (Fig. 2.6A). Substitution of A156 with either C or U resulted in a significant decrease in termination in the presence and absence of lysine (Fig. 2.6B). However, each variant exhibited a lysine-dependent increase in termination. The A156G variant exhibited a 3-fold increase in termination in the presence of lysine as compared to termination in the absence of lysine. The increase in termination when lysine was introduced was similar in the A156U variant, which exhibited a 2.5-fold increase (Fig. 2.6B). The A156C variant exhibited a 5-fold increase in the presence of lysine as compared to termination in the absence of lysine. It is noteworthy that the termination of the A156 variants was most efficient when A156 was substituted with a purine (A156G). This suggests that the heterocyclic identity at this position may be instrumental in the formation of the tertiary interaction. In addition, mutation at the A156 position did not cause a loss in lysine sensitivity. The retention of lysine sensitivity suggests that the formation of the helix 2-loop 4 tertiary interaction is not absolutely required for lysine-dependent termination.
Figure 2.6. *In vitro* transcription of A156 and A157 lysC leader RNA variants. (A) PAGE of single round transcription reactions of leader RNAs with mutations in A156 or A157 in the presence of 1.5 mM lysine. FL, full-length transcripts. T, terminated transcripts. (B) Bar graph of termination ratios for each transcription reaction. Open bars represent termination in the absence of lysine and filled bars represent termination in the presence of 1.5 mM lysine for each variant. Percent termination is the fraction of RNAs terminated relative to the total amount of RNA for each transcription. Reactions were performed in duplicate, and reproducibility was ±5%.
Mutation of A157 resulted in decreased termination in each of the variants analyzed. Unlike the A156 variants, each A157 variant exhibited reduced termination in the presence and absence of lysine, relative to wild type. Substitution of A157 with either C or U resulted in a 4.5-fold increase in termination when lysine was present as compared to termination in the absence of ligand. The largest change was exhibited by the A157G variant, which showed a 5-fold increase in termination in the presence of lysine. The increased termination in the presence of lysine indicates that mutation of either A156 or A157 does not significantly perturb lysine binding or the ability of the leader RNA to terminate transcription. Instead, the variants exhibited reduced termination as compared to wild type, suggesting that the mutations may have an effect on the structural transition of the RNA due to reduced stability of the global architecture.

2.4 Discussion

A common theme among riboswitch RNAs is the alternative folding of mutually exclusive structures, as conformational switching is imperative to the regulatory function of riboswitch RNAs. Helix 1 is a key component to the structural modulation. Although helix 1 is present in a large number of riboswitches the \textit{glmS} ribozyme and the \textit{SMK} riboswitch are two examples of riboswitches that do not contain this central helix. In both examples, the regulatory structure differs from that of a typical riboswitch. In the \textit{SMK} RNA, both the ligand-binding site and regulatory structure are located within the same helix, whereas in the \textit{glmS} ribozyme regulation is a result of an internal cleavage
reaction that occurs adjacent to the ligand-binding domain. Riboswitches that contain separate ligand-binding and regulatory structures often contain helix 1, formation of helix 1 is one of the first indications that the effector molecule is bound, and is coupled with the conformational change that modulates gene expression. Structural analysis has provided a wealth of information regarding the tertiary structures of riboswitch aptamer domains. In particular, high-resolution structures have provided insight into the global architecture of the RNA and molecular details of ligand recognition for at least 12 different riboswitch classes (Liberman and Wedekind, 2011). However, little is known about the ligand-dependent conformational changes that occur in response to the regulatory effector. Recent characterization of the SAM-responsive $S_{MK}$ riboswitch is one of few studies that offer an in depth analysis of the conformational changes of the aptamer and expression platform that occur upon ligand binding (Lu et al., 2008; Wilson et al., 2011). The architecture of the $S_{MK}$ RNA contains both the aptamer domain and expression platform within a single structure (Lu et al., 2008). In contrast, the $B. subtilis$ $lysC$ leader RNA is organized into two separate domains. In the presence of lysine, a series of intramolecular interactions within the $lysC$ leader RNA create a complex tertiary structure that results in the formation of an intrinsic terminator. Currently, it is not clear which intramolecular contacts are required for termination.

In this study, we focused our analysis on the identification of features of the L box leader RNA that are essential for lysine-dependent termination. Oligonucleotide-directed
RNase H cleavage was used to monitor the lysine-dependent formation of helix 1 in transcriptional and translationally active L box leader RNAs. Protection of helix 1 increased as the concentration of lysine was increased, which indicates that formation of helix 1 is lysine-dependent. This is consistent with previous *in vitro* studies of the *lysC* riboswitch that demonstrated that the leader RNA requires higher concentrations of AEC for termination similar to that promoted by lysine. This differential specificity was also observed in the RNase H cleavage assays in which protection of helix 1, at levels similar to that observed for lysine, required higher concentrations of AEC. Although similar concentrations of lysine and AEC were required for termination by the *E. coli* and *B. subtilis* riboswitches, the increased protection of the *B. subtilis* leader at lower lysine concentrations suggests that the *B. subtilis* leader RNA is more sensitive to lysine. The observed sensitivity might be an indication of the relative stabilities of the two AAT elements in the presence of lysine. The AAT of the *B. subtilis* leader RNA may be more stable in the presence of lower lysine concentrations, as compared to the *E. coli* AAT. Another possibility is that the *B. subtilis* leader RNA might have a higher lysine affinity and therefore can promote formation of helix 1 at lysine concentrations lower than those required by the *E. coli* L box leader RNA.

Previous studies of AEC resistance in *B. subtilis* and *E. coli* revealed that the resistance was the result of mutations in the *lysC* leader RNA (Vold *et al.*, 1975; Lu *et al.*, 1992; Patte *et al.*, 1998). This led to the assertion that AEC toxicity may be the
consequence of repression of \textit{lysC} synthesis under conditions that require lysine biosynthesis. Our analysis of the \textit{lysC} riboswitch from an AEC-resistant strain of \textit{E. coli} indicates that AEC toxicity is not the result of termination of \textit{lysC} expression by the lysine analog. Instead, toxicity is the result of AEC misincorporation into cellular proteins by LysRS2 (Ataide \textit{et al.}, 2007).

Conserved residues in helix 1 of the \textit{B. subtilis lysC} leader RNA were mutagenized to assess the importance of the sequence for lysine-dependent termination. Results of the mutational analysis uncovered sequences critical to the lysine-dependent structural transition. Variants with substitutions at U35-A198, that maintained base pairing, exhibited lysine-dependent termination but lower efficiencies were observed for some variants as compared to wild type. Substitution with a G-U pair also caused a reduction in the termination efficiency although base pairing at the position was maintained. This result suggests that the stability of the base pair at the 35-198 position is important for lysine-dependent termination. Substitution with the more stable A-U or G-C pairs resulted in termination efficiencies similar to those of wild type. The C35-G198 substitution, which presumably has the same stability as the G35-C198 variant, exhibited increased termination in the absence of lysine as compared to wild type. It is possible that the mutation caused instability in the AT. For example, the C35-G198 mutation in helix 1 resulted in a swap of A198-U248 for the less stable G198-U248 in the antiterminator. The increased stability of the terminator conformation may be the result
of antiterminator instability due to the G-U base pair although this substitution is not predicted to disrupt antiterminator formation. One explanation is that the G-U base pair reduces the ability of the antiterminator to compete for structure formation and causes the leader RNA to form the alternate termination conformation.

The A36-U197 base pair was less tolerant to mutation than the neighboring U35-A198 position. This is likely due to the location of the A36-U197 pair as it is closer in proximity to the lysine-binding pocket. Disruption of the base pair caused increased antitermination of leader RNA variants with the exception of C-A variant. Insertion of a C-A pair at the 36-197 position resulted in similar levels of termination in the presence and absence of lysine. The C-A mutation was predicted to disrupt the base pairing in both helix 1 and the antiterminator. However, the RNA was able to form a ligand-independent termination conformation. These results indicate that the sequence and base pairing interactions are important for the structural transition that causes termination of transcription. Substitution of the A-U pair with G-U, U-G, or U-A resulted in reduced termination as compared to wild type. The reduction in the termination efficiency observed in the A36-U197 variant, as compared to wild type, suggests that the stability of the base pair does not affect lysine-dependent termination. Similar to the U35-A198 position, alteration of the base pair composition of the helix had little effect on termination. In the presence and absence of lysine, termination of the G36-C197 variant was similar to wild type. This is in contrast to the C36-G197 variant where mutations
caused an increase in termination in the absence of lysine as compared to wild type. In addition, the U197G mutation results in substitution of a G197-A249 base pair in the antiterminator. Although non-canonical base pairing can occur with the given mutations, it is possible that A197G destabilizes the antiterminator and causes the leader RNA to preferentially terminate transcription in the absence of ligand.

Select helix 1 variants were tested in vivo to assess the behavior of the RNA under physiological conditions. Regulation of gene expression was consistent with data obtained from in vitro transcription. Expression of the U35-A198 and G35-C198 variants indicates that the substitution of a G-C at U35-A198 is more detrimental to gene expression than reversal of the base pair (A35-U198). The C-G variants at both positions exhibited low expression in the presence and absence of lysine, as compared to wild type. Leader RNA variants that exhibited lysine-dependent repression of gene expression in vitro exhibited reduced induction ratios as compared to wild type. Together, these results suggest that conserved helix 1 sequences at positions U35-A198 and A36-U197 play a key role in lysine-dependent termination. The stability of the individual structural elements is critical to the modulation of gene expression, as the leader RNA must quickly make a regulatory decision. The sequence of helix 1 is evolutionarily designed to maintain the appropriate balance between the two alternate competing structures, and alteration of this sequence results in changes in the structural stability that circumvent the lysine-dependent structural transition.
The long-range interaction between helix 2 and the helix 4 terminal loop was suggested to be critical to the L box leader RNA structural change (Blouin et al., 2011). Recent biochemical analysis suggested that the long-range interaction might play a key role in formation of the binding pocket (Serganov et al., 2009). Here we used mutational analysis of the B. subtilis lysC leader RNA to assess the sequence requirements for the helix 2-loop 4 interaction. In vitro transcription of variants with a disruption in the helix 2-loop 4 interaction demonstrated that these variants retain the ability to terminate transcription in the presence of lysine. The ability of the leader RNA to recognize and bind lysine suggests that the helix 2-loop 4 interaction is not directly involved in signal recognition. Instead, the mutations caused an alteration in termination efficiency where the leader RNA exhibited a reduction in termination, as compared to wild type. These results suggest that formation of the helix 2-loop 4 interaction is essential for termination efficiencies similar to wild type but is not required for lysine dependent termination. It is possible that mutation of a single conserved base only partially disrupted the interaction and the leader RNA is still able to make the long-range interaction with the remaining conserved base. This might account for the reduced termination efficiency that results from a weakened docking interaction.

The conserved sequences in helix 1 are crucial to the structural modulation of the lysC leader RNA in B. subtilis. Our analysis indicates that the formation of helix 1 in the L box riboswitch is more sensitive to lysine concentration when the leader RNA regulates expression at the transcriptional level as opposed to translation. The response to lysine
by the transcriptional L box riboswitch might be indicative of the requirement to make an immediate regulatory decision within the timeframe of transcription. Mutational and functional analysis of the lysC leader from B. subtilis reveals that formation of the termination conformation is largely dependent upon helix 1. Changes to the base pair ratio of the helix did not significantly affect lysine-dependent termination. However, substitution of either base pair with a C-G pair resulted in an alteration of RNA stability in which the termination conformation was more stable in the absence of ligand. Alteration of the helix 2-loop 4 long-range tertiary interaction suggests that this interaction plays an instrumental role in stabilization of the global structure of the RNA but is not involved in stabilization of the lysine-binding pocket as lysine-dependent termination by the variants was not perturbed. This mutational analysis provides an additional example of the simple composition yet complex function of riboswitch RNAs wherein modulation of the RNA structure is finely tuned by the sequence. It is intriguing that single mutation can alter the functionality of the riboswitch. Changes in a single base pair in helix 1 preclude the structural transition, which results in an altered folding trajectory. Further analysis of riboswitch sequence as it relates to the folding of the structure will provide an alternative perspective on the features necessary for regulation of gene expression.
CHAPTER 3

IN VITRO EVOLUTION OF VARIANT L BOX LEADER RNAs

3.1 Introduction

Currently, there are 20 known classes of riboswitches that regulate gene expression in response to at least 15 small molecules (Mairal et al., 2008). For this mechanism to be effective, the riboswitch must respond only to the appropriate molecule. The heterogeneous environment of the cell makes this a daunting task, as the riboswitch must selectively recognize the cognate ligand and bind with an affinity appropriate for the specific ligand. Thus, ligand affinity and specificity are of great importance to this regulatory mechanism. The ligand-binding domain is a highly conserved structural feature among all riboswitch RNAs, responsible for the recognition and binding of the effector molecule. Recognition of the signal causes conformational changes within the RNA that result in the formation of a regulatory structure that can activate or repress transcription (intrinsic terminator), prevent or promote translation (Shine-Dalgarno sequestering helix), or alter mRNA processing.

Riboswitches are known to bind the cognate ligand with a physiologically relevant affinity and discriminate against similarly structured molecules. These RNAs have an exquisite selectivity for the natural effector and can recognize molecular changes
in the ligand such as the substitution of a carbon for sulfur. The S box RNA is one example that illustrates the discriminatory nature of riboswitches. In *B. subtilis*, the S box is located upstream of several genes involved in SAM biosynthesis. This leader RNA modulates gene expression based on the intracellular SAM concentrations. Studies of the S box regulon indicated that riboswitches from the same class exhibit differential affinities for SAM (Tomsic *et al.*, 2008). Although the sequence and structure of the S box ligand-binding domain is highly conserved, each riboswitch is calibrated to respond to a specific SAM concentration. The SAM affinity of the *B. subtilis yitJ* S box leader is estimated in the low nanomolar range ($K_D \sim 20$ nM) (McDaniel *et al.*, 2003; Winkler *et al.*, 2003). The SAM affinity of the *yxjG* and *yoaD* S box leader RNAs was estimated around 400 and 230 nM, respectively, a 10- to 20-fold decrease in affinity as compared to the *yitJ* leader RNA (Tomsic *et al.*, 2008). The weakest affinity was observed in the *yusC* leader RNA that exhibited a $K_D$ value of 3.5 μM. Overall, a 250-fold range in SAM affinities was observed among the S box leader RNAs, which suggests that these differences are likely due to sequences and structural features that are unique to each riboswitch. In addition to the differential SAM affinity, S box leader RNAs can discriminate between SAM and SAH with affinities that differ by as much as three orders of magnitude (McDaniel *et al.*, 2003; Wang and Breaker, 2008). Although SAM and SAH only differ by a methyl group and the molecular charge of the sulfur, analysis of the specificity determinants of the S box riboswitch indicate that both features affect the binding affinity (Lu *et al.*, 2010; Ham *et al.*, 2011).
The adenine and guanine riboswitches also exhibit a high affinity for the effector. Despite the sequence and structural similarities between the purine riboswitches, both RNAs recognize the cognate ligand with high specificity while discriminating against similar molecules (Kim and Breaker, 2008). The guanine riboswitch, upstream of the \textit{xpt-pbuX} operon, binds guanine with high affinity ($K_D \approx 5$ nM), and binds adenine with an approximate $K_D$ of \approx 300 nM (Mandal et al., 2003). The S box and purine riboswitches illustrate the high affinity and selectivity harnessed within the riboswitch ligand-binding domain. These features are required to modulate gene expression in response to intracellular metabolite concentrations.

The lysine-responsive riboswitch is composed of a ligand-binding domain that binds lysine at a 1:1 ratio (Sudarsan et al., 2003). \textit{In vitro} transcription termination assays demonstrated that the L box leader RNA is specific for lysine and can discriminate against closely related lysine analogs (Grundy et al., 2003). The concentration required for half-maximal termination (Trm$_{1/2}$) of the \textit{B. subtilis} lysC leader was \approx 170 μM in the presence of lysine and 800 μM in the presence of AEC (section 4.3.1). The differential sensitivity that the leader RNA exhibits for lysine, as compared to lysine analogs, is also an indication of the ligand affinity (section 4.3.1). Analysis of the S box leader RNAs illustrates a correlation between the SAM affinity and the concentration required for termination \textit{in vitro} (Tomsic et al., 2008). Studies of the lysine affinity of a minimal \textit{B. subtilis} lysC leader RNA construct indicated that the RNA binds lysine with an approximate $K_D$ of 1 μM (Sudarsan et al., 2003; Blount et al., 2007). This affinity is
significantly lower than other riboswitches that often bind the cognate ligand with a nanomolar range affinity. Analysis of the full-length *lysC* leader RNA construct that included the expression platform resulted in an approximate $K_D$ of $\sim 500 \mu M$ (Sudarsan *et al.*, 2003; Blount *et al.*, 2007). The increase in $K_D$ when the expression platform was present suggests that additional structures downstream of the aptamer domain contribute to the overall lysine affinity. The change in affinity of the full-length construct as compared to the ligand-binding domain might be due to the competing structures but may also more accurately represent the physiological binding affinity.

To date, few studies have focused on examining the sequence and structural features of the L box riboswitch that contribute to lysine affinity. *In vitro* selection by SELEX (Systematic Evolution of Ligands by EXponential enrichment) is a synthetic biology technique that allows *in vitro* evolution of nucleic acids based on a specific functionality (Ellington and Szostak, 1990; Tuerk and Gold, 1990). In this method, successive rounds of enrichment and amplification of a large variant pool is used to identify unique molecules harboring a specific function. One of the first reports of *in vitro* evolution resulted in the identification of a high-affinity theophylline RNA aptamer (Jenison *et al.*, 1994). The aptamer bound theophylline with a $K_D$ of approximately 0.1 $\mu M$. This is 10,000-fold higher than the affinity for caffeine, a theophylline analog. The SELEX method used for identification of the theophylline aptamer can be modified for various biological applications, particularly for the development of nucleic acid aptamers.
Currently, limited information is available about the sequence and structural features that contribute to the lysine affinity of the L box riboswitch. It is our goal to identify sequence and structural features of the L box leader RNA that constitute the molecular basis for lysine affinity. SELEX is ideal for this aim as riboswitches are naturally occurring aptamers specific for small organic molecules. Thus, SELEX provides an avenue for the identification of features that contribute to ligand affinity through the selection of variant molecules with altered binding properties. In this study, we used SELEX to identify features of the \textit{B. subtilis lysC} leader RNA that are responsible for lysine affinity. \textit{In vitro} evolution of a variant pool with the ability to respond to lysine at concentrations lower than wild type identified sequences and structures within the \textit{lysC} leader RNA that are important for lysine binding and regulation of gene expression.

3.2 Materials and Methods

3.2.1 Construction of DNA templates

The DNA templates for T7 RNAP transcription were generated as described above (section 2.2.1). For the lysine binding assays, the wild-type \textit{lysC} fragment was generated using complementary pairs of overlapping DNA oligonucleotides (Table 2.1), and the DNA source for PCR amplification using \textit{Taq} DNA polymerase (Invitrogen, Carlsbad, CA) (McDaniel \textit{et al.}, 2005; Yousef \textit{et al.}, 2005). DNA oligonucleotides \textit{lysCT7US} and \textit{lysCBsp3’hlx1} (Table 2.1) were used as 5’ and 3’ primers, respectively.
The resulting fragment contained the T7 RNAP promoter fused to position +18 of the lysC leader sequence and ended at position 234. The PCR product was purified using a QIAquick PCR clean up kit (Qiagen, Chatsworth, CA) and sequenced by Genewiz (South Plainfield, NJ) to confirm accuracy in template construction and sequence.

DNA templates for the RNase H protection assays were generated using overlapping complementary DNA oligonucleotide pairs as described above (section 2.2.1). Variants isolated from the SELEX pool were constructed using site-directed mutagenesis of the wild-type leader template. DNA oligonucleotide pairs that contained the desired mutations were substituted for the corresponding wild-type pair in the kinase reaction. The variant and wild-type pairs were ligated, amplified by PCR, purified, and sequenced.

Templates used for in vitro transcription were constructed as described above. Briefly, a transcriptional fusion plasmid, derived from pFG328, that contained the wild-type lysC leader sequence downstream of the glyQS promoter (section 2.2.1) was used as the template for site-directed mutagenesis. DNA oligonucleotide pairs with mutations, corresponding to the variants isolated by SELEX, were used as primers for PCR amplification of the pFG328-lysC plasmid using Pfu DNA polymerase (Stratagene, La Jolla, CA). The PCR reactions were digested with DpnI (NEB, Beverly, MA) to remove the residual starting template. After digestion, the plasmids were introduced into XL-2 blue ultracompetent cells (Stratagene, La Jolla, CA) per the manufacturer’s instructions.
Cells that contained the desired plasmid were isolated and the plasmid DNA was purified and sequenced as previously described (section 2.2.1). The final template used for in vitro transcription contained the glyQS promoter sequence upstream of the variant lysC leader sequence.

Table 3.1. DNA oligonucleotides for circularly permutated lysC leader RNA synthesis

<table>
<thead>
<tr>
<th>Primer</th>
<th>Position</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>lysCT7cpRNA2</td>
<td>T7 promoter 185-35</td>
<td>TAAATCGACTCACTATAGTTTCTTGGAGGGCTATCTGCAAAAAAAGAGAT</td>
</tr>
<tr>
<td>lysCT7cpRNA2RC</td>
<td>T7 promoter 185-30</td>
<td>CTTTTTTTTTGCAGATAGCCCTCCAAGAAACTATAGTGAGTGGATTTA</td>
</tr>
<tr>
<td>lysCcpRNA casA</td>
<td>36-59</td>
<td>AGAGGTCGGAACCTCAAGAGATAGG</td>
</tr>
<tr>
<td>lysCcpRNA casA RC</td>
<td>30-54RC</td>
<td>TCTTGAAGTTCGACCTCTATCCTC</td>
</tr>
<tr>
<td>lysC CasB wt</td>
<td>60-104</td>
<td>CTTTGGAGAAAGATGGATCTGTAAGAAAAGGGCTGAAGGAGAGC</td>
</tr>
<tr>
<td>lysC CasB wt RC</td>
<td>55-99</td>
<td>CTTTCAAGCTTTTTTACAGAATCCATCTCTCTTCACCAAGGATAC</td>
</tr>
<tr>
<td>lysCcpRNA casD</td>
<td>105-154</td>
<td>GCCTCAGGAGCAAAATAAACCACATCGGTATTATTCCTGGCCCGTGCAT</td>
</tr>
<tr>
<td>lysCcpRNA casD RC</td>
<td>100-154rc</td>
<td>AATGCACGGCCAGCAAAATACCGATGGGGTTTATTCTCGCGACGCTCC</td>
</tr>
<tr>
<td>lysCRNH5</td>
<td>256-265</td>
<td>AAAAGCAATGA</td>
</tr>
<tr>
<td>lysC term</td>
<td>209-258</td>
<td>TGAGAGGAAATACCTCTATTGCTATTATATAATCAGATGGGATTTATTGCTG</td>
</tr>
<tr>
<td>lysC term ext</td>
<td>243-266+23 rand bp</td>
<td>GTTTGACACAAGGTACAGCATTATCTGGAAAAAAGAGAT</td>
</tr>
</tbody>
</table>

To monitor the lysine–dependent structural changes of the lysC leader RNA, we used a site-specifically incorporated fluorescent probe at A156. The template was divided in half to allow for the repositioning of the 5’ and 3’ termini. The 5’ half of the RNA (155-183 Fig. 3.3) was commercially synthesized for 2-AP incorporation (Dharmacon, Chicago, IL). The 3’ half contained positions 184-154 and was circularly permuted at positions 203 and 30 (Fig. 3.3). To construct the DNA template for the 3’ half of the RNA, which corresponds to positions 184-154, the phage T7 promoter sequence was fused to position 184. A 9-base adenosine linker was introduced between positions 203 and 30 (Table 3.1) to circularize the template. The remaining portion of the
sequence was constructed using overlapping DNA oligonucleotide pairs (Table 3.1) that were kinased and ligated as described above (section 2.2.1). The resulting DNA template was amplified using the flanking 5’ and 3’ DNA oligonucleotides as primers for PCR by Taq DNA polymerase (Invitrogen, Carlsbad, CA). The final template was purified and sequenced as described above.

3.2.2 RNA synthesis and purification

The RNAs for the 3’ half of the bipartite RNA and the lysine-binding assay were synthesized using a MEGAshortscript T7 RNAP transcription kit (Ambion, Austin, TX). The transcripts were excised from a polyacrylamide gel after denaturing PAGE and passively eluted into 1X transcription buffer (20 mM Tris-HCl pH 8.0, 20 mM NaCl, 10 mM MgCl$_2$ and 0.1 mM EDTA). The RNA products were purified by phenol-chloroform extraction and ethanol precipitation. The concentration of the RNA was determined using a NanoDrop spectrophotometer (Thermo Fisher Scientific, Wilmington, DE). The 5’ half of the bipartite RNA corresponding to positions 155-183 of the lysC leader RNA was commercially synthesized with a 2-AP substitution at position A156 (Dharmacon, Chicago, IL).

3.2.3 Lysine-binding assay

T7 RNAP transcribed lysC leader RNAs corresponding to positions 18-234 were diluted to a final concentration of 1 μM in 1X transcription buffer and denatured at 65°C
for 5 min. The RNA was slow-cooled to 40°C for refolding and L-[4, 5-3H(N)] lysine (110 Ci/mmol [4.07 TBq/mmol], Perkin Elmer) was added to achieve the indicated final concentrations. The reactions were incubated at room temperature for 5 min to allow lysine binding. Each reaction was transferred into a Nanosep microconcentrator filtration device (Pall Corporation) and centrifuged at 14,000 × g for 2 min. The filters were washed twice with 1X transcription buffer to remove any residual unbound lysine. The products that remained on the filter were extracted, mixed with Ultima Gold scintillation fluid (Packard BioScience, Meriden, CT), and the radioactivity of the mixture was measured using a Packard Tri-Carb 2100TR liquid scintillation counter. Each reaction was performed in duplicate and reactions without RNA were included as negative controls.

3.2.4 Fluorescence spectroscopy

The RNA structural changes induced by ligand binding were monitored using fluorescence spectroscopy. The aptamer domain of the lysC leader RNA was designed in two halves and circularly permutated for the site-specific incorporation of the fluorescent adenine analog 2-aminopurine (2-AP) at position A156 (Dharmacon, Chicago, IL) (section 3.2.1). The two RNA halves were mixed (at a final concentration of 1 μM) in 1X transcription buffer and incubated at 65°C for 5 min. The RNA was slow-cooled to room temperature for refolding. Lysine was added to the indicated final concentrations and the mixture was incubated 30 min at room temperature and transferred to a quartz
cuvette. 2-AP fluorescence was monitored in the presence and absence of lysine using a FlouroLog-3 Spectrofluorometer (Horiba Scientific, Edison, NJ). The reaction was exposed to 310 nm for excitation and 2-AP emission was monitored between 330 and 420 nm wavelengths. The change in fluorescence was determined by the 2-AP emission at 375 nm in the presence as compared to in the absence of lysine.

3.2.5 *In vitro* selection using SELEX

The *lysC* leader template (450 bp), which contained sequences 83 bp upstream of the transcription start-site and 95 bp downstream of the terminator, was mutagenized to generate a variant DNA pool. For mutagenesis, MnCl$_2$ (10 mM) was added to the PCR mixture during template amplification (D. Frisby, unpublished results). The mutagenesis was monitored by cloning and sequencing of samples from the PCR products. The DNA pool selected as the starting template for the *in vitro* selection contained approximately 1 to 3 base changes per molecule. After mutagenesis, an additional PCR was performed to produce a template that contained positions 18-234 of the *lysC* leader region. The 5’ primer for PCR contained the phage T7 promoter sequence fused to the +18 position relative to the *lysC* transcription start site. The 3’ oligonucleotide hybridized internally and ended at position 234, 3 nt prior to the 5’ position of the transcription terminator. The products of the PCR were purified as described above.

The starting RNA pool for *in vitro* selection was generated by T7 RNAP transcription of DNA (1.2 pMol) from the mutagenic PCR. Transcription was carried out
using a MEGAscript T7 RNAP transcription kit (Ambion, Austin, TX) in the presence of 0.5 mM lysine for 30 min at 37°C. After transcription, a DNA oligonucleotide (47 µM) complementary to positions 193-204 of the *B. subtilis* *lysC* leader RNA was added and the reaction was incubated for 5 min at 37°C. RNase H (0.45 U; Ambion, Austin, TX) was added and the reaction was incubated for an additional 10 min at 37°C. The reactions were stopped by phenol-chloroform extraction and the resulting RNA products were resolved by denaturing PAGE on a 6% gel. Transcripts that were protected from RNase H cleavage were excised from the gel and passively eluted overnight in RNase-free 1X transcription buffer. The resulting RNAs were purified further by phenol-chloroform extraction and ethanol precipitation. After purification, the RNA mixture was digested with DNase I using a DNAeasy kit (Ambion, Austin, TX), per the manufacturer’s instructions, to remove any residual DNA template. cDNA was synthesized from the RNAs using a ThermoScript RT-PCR kit (Invitrogen, Carlsbad, CA). Briefly, a DNA oligonucleotide complementary to positions 201-234 was used as a primer for cDNA synthesis. The DNA oligonucleotide and RNA were incubated at 65°C for 5 min for hybridization. The reaction was added to a cDNA synthesis mixture that contained reverse transcriptase and incubated for 60 min at 58°C. To stop the reaction, the mixture was heated to 85°C for 5 min. Next, RNase H (2 U mL⁻¹) was added to remove the starting RNA template and the remaining cDNA was amplified by PCR using *Taq* DNA polymerase (Invitrogen, Carlsbad, CA). This DNA served as the starting template for the T7 RNAP transcription for the next round of
selection. The variant RNA pool was initially subjected to nine rounds of \textit{in vitro} selection. The DNA template from the ninth round of selection was cloned and sequenced (Genewiz, South Plainfield, NJ). Analysis of the variants from several rounds of \textit{in vitro} selection required that of the selection procedure be modified to include an additional selective pressure. This additional step was introduced into the SELEX scheme to select for RNAs that form the antiterminator in the absence of lysine as the previous selection scheme only required the leader RNAs to form the terminator in the presence of lysine. This additional step will eliminate molecules from the pool that are unable to switch between the two RNA conformations in the presence and absence of lysine.

In a PCR of the variant pool, we used two DNA oligonucleotides, the reverse complement of positions 209-266, to extend the DNA template to include the antiterminator element. This template was then used for transcription of a variant RNA pool. The RNA was transcribed in the absence of lysine for 30 min at 37°C. After transcription, a DNA oligonucleotide (47 µM) complementary to positions 256-265 of the \textit{B. subtilis lysC} leader RNA was added and incubated for 5 min at 37°C. RNase H (0.45 U; Ambion, Austin, TX) was added and the reaction was incubated for 10 min at 37°C. The reactions were stopped by phenol-chloroform extraction and the resulting RNA products were resolved by denaturing PAGE on a 6% gel. Transcripts that were cleaved by RNase H were excised from the gel and passively eluted overnight in RNase-free 1X transcription buffer. The leader RNAs were reverse transcribed and the cDNA
was used as the template for the next round of selection, as described above. Three additional rounds of selection performed and the resulting templates were cloned as sequenced.

### 3.2.6 RNase H cleavage assay

RNase H cleavage was carried out as previously described (section 2.2.3), with a few modifications. The DNA templates for T7 RNAP transcription were generated by PCR of templates isolated from *in vitro* selection (section 3.2.5). Transcription of the DNA was carried out using a MEGAscript T7 RNA polymerase kit (Ambion, Austin, TX) in the presence of 0.5 μM [α-32P]-UTP (GE Healthcare; 800 Ci/mmol [30 TBq/mmol]) for radiolabelling. RNAs were transcribed in the presence of lysine, at the concentrations indicated, for 30 min at 37°C. After transcription, a DNA oligonucleotide (47 μM) complementary to positions 193-204 of the *B. subtilis lysC* leader RNA was added and incubated for 5 min at 37°C. RNase H (0.45 U; Ambion, Austin, TX) was then added to the reaction and incubated 10 min at 37°C. The reactions were stopped by phenol-chloroform extraction. The resulting RNA products were resolved by denaturing PAGE and visualized by PhosphorImager analysis (Molecular Dynamics). Reactions were performed in duplicate and variability was ±10%.
3.2.7 *In vitro* transcription of SELEX isolated variants

Single round transcription of the variant *lysC* leader templates isolated from the *in vitro* selection was performed using the methodology as previously described (section 2.2.4). Briefly, single-round transcription was carried out using *B. subtilis* RNAP. The wild type and variant *lysC* leader RNAs (see section 3.2.1) were transcribed in the presence or absence of lysine at the indicated final concentrations. Lysine-dependent termination was monitored and the percent of terminated transcripts relative to the total RNA transcribed was determined. The reactions were performed in duplicate, and reproducibility was ±5%.

3.2.8 *lysC-lacZ* transcriptional fusions and β-galactosidase assays

Site-directed mutagenesis of a *lysC-lacZ* transcriptional fusion construct was performed using the vector pFG328 that contained the full-length wild-type leader RNA (see section 2.2.2). The mutations identified in variants isolated from the SELEX pool were introduced into the wild-type transcriptional fusion. A single copy of each construct containing the variant *lysC* leader sequence was integrated into the chromosome of strain BR151 (*lys-3 metB10 trpC2*), as described by Grundy *et al.* (2003). Cell culture and β-galactosidase measurements were carried out as described above (see section 2.2.5). Experiments were carried out in duplicate and variability was ±15%. 
3.3 Results

3.3.1 Analysis of lysine binding by the B. subtilis lysC leader RNA

To determine the optimal lysine concentration for in vitro selection using SELEX, we used size exclusion filtration to estimate the binding affinity of the lysC leader RNA. Previous analysis of the S box, S MK, and TPP riboswitches were successful in $K_D$ determinations using this method (McDaniel et al., 2003; Fuchs et al., 2006; Ontiveros-Palacios et al., 2008). In this assay, Nanosep filters with 10 kDa molecular weight cut-off (MWCO) were used to determine the binding affinity of the lysC leader RNAs from both B. subtilis and E. coli. The S box leader RNA found upstream of the yitJ gene, which does not bind lysine, was included as a negative control. T7 RNAP-transcribed RNAs were denatured and allowed to re-fold in the presence of $^3$H-lysine. Centrifugation of the reaction mixture in the filtration device allowed the separation of free lysine molecules from those bound by the RNA, which were retained on the filter. Filtration of the B. subtilis and E. coli lysC leader RNAs in the presence of lysine resulted in binding of only 0.6 and 1.2% of the lysine added to each reaction (Fig. 3.1). These values are slightly higher than the amount of residual lysine on the filter in the absence of RNA (0.4%). The S box yitJ leader RNA is specific for SAM and is not predicted to bind lysine, as previous results indicate that transcription in the presence of lysine had no effect on S box leader termination. The yitJ construct exhibited 0.8% binding, which
suggests that the binding observed for the L box leader RNAs is likely to be non-specific, possibly due to interactions between lysine and the RNA backbone.

Several modifications of the procedure were employed to determine if this assay was suitable for monitoring lysine binding. The goal for these experiments was to establish a lysine-binding assay suitable for the analysis of leader variants with altered lysine affinities. Previous in vitro studies of the lysC leader RNA indicated that the leader RNA responds to higher ligand concentration as compared to other riboswitches (Sudarsan et al., 2003; Blount et al., 2007). Thus, we increased the concentration of both lysine and the leader RNA in independent experiments. Although the lysine concentrations were increased, there was no increase in lysine binding (data not shown). Reactions with increased RNA concentrations, as compared to previous experiments, resulted in increased non-specific binding as indicated by the B. subtilis yitJ leader RNA control.
Figure 3.1. Lysine binding to the *B. subtilis* and *E. coli* lysC leader RNAs. $^3$H-lysine (6 μM) and leader RNAs (1 μM) transcribed by T7 RNAP were incubated. The mixtures were washed with 1X transcription buffer and concentrated using 10 kDa MWCO Nanosep filters. The lysine retained is expressed relative to the amount of lysine added to the mixture. The lysine retained on the filter in the absence of RNA was 0.4%.

3.3.2 Design of a circularly permutated bipartite RNA with a 2-AP substitution

In an alternative approach to measure the lysine binding affinity, we designed a circularly-permutated bipartite RNA that consisted of the *lysC* leader ligand-binding domain, corresponding to positions 30-234 of the *lysC* leader sequence. An RNA molecule of this type is suitable for the site-specific incorporation of a fluorescent molecule to allow monitoring of lysine-dependent changes in fluorescence emission. First, we introduced a 9-nt adenosine linker to circularize the leader RNA and enclose
helix 1. The introduction of the adenosine linker was crucial as it allowed us to reposition the 5' and 3' termini. The new termini were introduced between positions U154 (3’) and G155 (5’) near the terminal loop of helix 4 (Fig. 3.2). This circularly permutated $lysC$ variant was tested in RNase H cleavage assays to confirm that addition of the adenosine linker did not significantly affect formation of the tertiary structure. In the presence of lysine, the circularly permutated leader RNA exhibited protection of helix 1, while cleavage of helix 1 was observed in the absence of lysine (data not shown). We also added an additional break near the terminal loop of the helix 5 to divide the RNA in half. This nick in the RNA backbone was located between U184 (3’) and U185 (5’) (Fig. 3.2). The larger 5’ half of the leader RNA consisted of nucleotides 183-153 and included the 9-nt linker between positions 203 and 30. The 3’ half of the RNA, positions 155-184, was designed such that the fragment could be commercially synthesized and the fluorescent 2-AP molecule could be incorporated at residue A156.
Figure 3.2. Bipartite RNA used to monitor lysine binding in fluorescence assays. The \textit{l}ys\textit{C} leader RNA ligand-binding domain was constructed in two halves. The 5' half is shown in bold letters corresponding to positions 155-184 with a 2-AP substitution at A156. The 3' half of the RNA contains positions 185-154 with a 9-nucleotide adenosine linker between residues 30 and 204. The breaks in the RNA backbone are located at U154-G155 and U184-U185. Positions A156 and A157 are predicted to interact with helix 2 when lysine is bound. Nucleotides shown in red are 100% conserved, blue >80%, and green exhibit extensive covariation. Helices are numbered and AAT represents the anti-antiterminator region.
A few studies have identified regions of the lysC leader RNA that undergo structural modulation in the presence of lysine. In particular, the single stranded region between helices 2 and 3 was found to be highly sensitive to chemical probing in the absence of lysine but exhibited reduced reactivity when lysine was present (Garst et al., 2008). Mutation of the same residues in the single stranded junction between helices 2 and 3 also resulted in a loss of lysine-dependent termination (Grundy et al., 2003). Crystal structure analysis of the L box leader RNA revealed that the conserved single-stranded core of the molecule forms the binding pocket and directly contacts lysine (Garst et al., 2008; Serganov et al., 2008). Other regions of the RNA that undergo structural modulation in the presence of lysine include a docking interaction between helices 2 and 4. We postulated that this docking interaction is lysine-dependent and causes significant structural changes in the universally conserved A156 and A157 residues in the terminal loop of helix 4. Therefore, position A156 was selected for the site-specific incorporation of 2-AP. 2-AP is a fluorescent adenine analog often used as a tool to monitor local changes in nucleic acid structure (Ward et al., 1969; Soulière et al., 2011). 2-AP is suitable for the analysis of structural transitions because it can be site-specifically incorporated and can participate in Watson-Crick base pairing interactions with uracil or wobble interactions with cytosine. Incorporation of a 2-AP molecule is predicted to cause minimal perturbation to the RNA structure. 2-AP absorbs at excitation wavelengths between 300 and 330 nm and emission can be monitored between 330 and 420 nm with the maximum excitation wavelength of 375 nm (Ward et al., 1969). The 2-
AP molecule is extremely sensitive to changes in the local environment, thus, changes in RNA structure can be monitored as a function of changes in 2-AP fluorescence (Ward et al., 1969; Soulière et al., 2011).

3.3.3 Analysis of lysine binding using spectrofluorimetry

To determine if this assay was appropriate for the estimation of lysine affinity, we used a bipartite lysC leader RNA harboring a 2-AP molecule to monitor changes in the RNA structure as a function of lysine concentration. We first determined the emission spectra of the 3’ half of the RNA that contained the 2-AP molecule, the bipartite RNA, or the bipartite RNA in the presence of lysine. The fluorescent emission of the 2-AP molecule alone was partially quenched when the remaining half of the leader RNA was added (Figure 3.3A). This suggested that the reduction in emission was due to hybridization between the two molecules that caused changes to the local 2-AP environment. To ensure that the fluorescent signal was from the 2-AP molecule, reaction mixtures that contained only water, 1X transcription buffer, or lysine, were also monitored and little fluorescence was observed (Fig. 3.3a; data not shown). When lysine was added to the mixture, the 2-AP molecule exhibited additional quenching, which suggests that the terminal loop of helix 4 becomes more structured in the presence of lysine (Fig. 3.3B). This is consistent with the docking interaction in helix 4 that was observed in the crystal structure of the lysine-bound anti-antiterminator domain (Serganov et al., 2009). To determine if the fluorescent quenching was concentration-
dependent, lysine was added to reaction mixtures in 2-fold increasing increments (Fig. 3.3C). The increase in lysine concentration also resulted in fluorescent quenching; however, the change in fluorescence was not consistent with concentration-dependent quenching of fluorescence as the lysine concentration was increased. We expected an incremental increase in fluorescence quenching as the lysine concentrations were increased. At 375 nm, the wavelength of maximum 2-AP emission, a 6% increase in quenching was observed in the presence of 54 uM lysine (87%) as compared to 13.8 mM lysine (92%) (Fig. 3.3D). Several modifications to the assay were performed, such as decreasing the lysine concentration, reducing the RNA concentration, the addition of Mg$^{2+}$, and changes to the reaction temperature. None of the methods employed resulted in an increase in the dynamic range of the emission spectra of the 2-AP molecule. The small change in fluorescence suggested that this method was not suitable for the estimation of lysine affinity as titration of lysine was not effective. Our initial goal for this work was to identify variant lys$C$ leader RNAs with a lysine affinity lower than wild type. In order to identify those RNAs with increased lysine sensitivity we must have a consistent and reproducible method to monitor the lysine affinity. The experiments utilized here to determine the lysine affinity were unreliable and we were subsequently required to approach the in vitro selection using an alternate strategy.
Figure 3.3. Fluorescent emission of 2-AP labeled bipartite leader RNA. Fluorescent emission of the 2-AP molecule was monitored between 330-420 nm after excitation at 310 nm. A) The 5’ half of the RNA harboring the 2-AP molecule exhibited the maximum emission spectra. Addition of the 3’ half of the bipartite RNA resulted in partial fluorescent quenching. Further quenching was observed upon the addition of lysine. B) To determine if quenching was concentration-dependent, lysine was added at each of the indicated concentrations and fluorescent quenching was monitored. C) Fluorescent quenching of the bipartite RNA over a wide range of lysine concentrations (13.8 mM - 54 uM). D) Percent quenching is plotted as a function of lysine concentration. The percent quenched is the change in fluorescence emission at 375 nm relative to the fluorescence at in the absence of lysine.
3.3.4 Design of SELEX selection scheme

Although we were unable to measure the lysine affinity using previously established methods we used \textit{in vitro} evolution to uncover features of the \textit{lysC} leader RNA involved in the lysine affinity. The SELEX method was selected based on the adaptability to the riboswitch system and previous reports of \textit{in vitro} evolution that identified high affinity RNA aptamers specific for small molecules (Ellington and Szostak, 1990). The natural aptamer domain of the \textit{lysC} riboswitch offers a pre-existing RNA scaffold for \textit{in vitro} evolution of L box riboswitches. Therefore, a SELEX scheme was designed to select for variants of the \textit{B. subtilis lysC} leader RNA that bind lysine at concentrations lower than wild type.

In most SELEX methods the typical starting nucleic acid pool is based on the length of the target sequence and often composed of a randomized library with $4^n$ (n=length of sequence) number of sequences for the identification of a unique aptamer. The \textit{lysC} riboswitch is a highly conserved RNA scaffold that relies on the global structure of the RNA to bind lysine, thus, a completely randomized template pool was not suitable for our purposes. Instead, we wanted to preserve the tertiary architecture of the RNA and use SELEX to identify variants of the leader RNA with altered ligand-binding properties. For a variant pool with an average mutation of 1 per molecule the starting pool required less than 1,000 molecules for complete mutagenic coverage. Here we used mutagenic PCR to randomly mutate the 204 bp \textit{lysC} leader fragment that contained the lysine-
binding pocket and the anti-antiterminator region of the leader RNA. Several PCR reactions were performed and the template pool selected contained an average of 1 to 3 mutations per molecule. To ensure that the tertiary architecture of the leader RNA was maintained throughout the selection, DNA oligonucleotide primers were designed to protect the 5’ and 3’ sequences of helix 1 and used at each step of amplification. This step was included to eliminate mutations known to disrupt lysine binding and formation of the tertiary structure.
Figure 3.4. *In vitro selection scheme*. Wild-type *lysC* leader DNA was used as a template in mutagenic PCR to create a pool of variant templates. The pool was then transcribed in the presence of lysine. A DNA oligonucleotide complementary to the 5’ side of helix 1 was allowed to hybridize and RNAs were challenged with RNase H. Denaturing PAGE was used to separate the RNAs. The full-length product that was protected from cleavage was further purified. Next, reverse transcriptase was used to make a cDNA copy of the full-length RNAs. PCR was used to amplify the cDNA and enrich the DNA pool used for the next round of selection.

Our preliminary studies of helix 1 formation, using the RNase H assay, indicated that formation of helix 1 was dependent upon the lysine concentration (section 2.3.1). As
the concentration of lysine was increased, the wild type \textit{lysC} leader RNA exhibited increased protection of helix 1, whereas in the absence of lysine, RNase H cleaved the leader RNA. This method allows us to differentiate between RNAs that form the termination or readthrough conformation in response to the lysine concentration and is a reliable indication of lysine affinity based on the lysine-dependent helix 1 formation. This method was incorporated into our SELEX procedure to select for RNAs that form helix 1 in the presence of lysine at concentrations lower than the wild-type RNA (Fig. 3.4). The variant template pool generated using mutagenic PCR was the starting template for T7 RNAP transcription. After transcription, RNAs unable to form helix 1 were cleaved by RNase H and eliminated from the pool. The remaining RNAs were reverse transcribed and the cDNA was amplified by RT-PCR. Variants from iterative rounds of selection were isolated, cloned, and sequenced (Fig. 3.4).

3.3.5 Identification and characterization of SELEX variants

\textit{In vitro} selection of a DNA template pool containing 1-3 mutations per molecule was initiated, and the pool was subjected to nine successive rounds of selection. After each round, the DNA pool was transcribed and tested in an RNase H assay to monitor helix 1 formation as a function of the lysine concentration. Analysis of the variants that were cloned and sequenced from the selection revealed a group of isolates with at least 5-7 single nucleotide substitutions per molecule. The average number of changes per molecule indicated that majority of the variants had more mutations than originally
anticipated. Although the starting pool only contained 1-3 changes per molecule, it is possible that the subsequent amplification steps introduced additional mutations. To determine how many rounds of selection were required to maintain a variant pool with an average of 1-3 base changes per molecule, we performed RNase H analysis of each round, which revealed that the overall protection of the RNA pool slightly increased after each round of selection. However, a dramatic shift in helix 1 protection was observed after the third round of selection. Protection of the leader RNAs from the third round of selection was ~60-70% in the presence of 0.5 mM lysine. RNase H protection of subsequent rounds, at the same lysine concentration, exhibited an increase in protection to ~85-95%. Therefore, the DNA pool that was subjected to three rounds of selection was chosen for cloning and sequencing.

In total, 25 variants were isolated and sequenced and each contained between 1-3 mutations. The mutations were located at various positions within helices 2-5 and only a few were observed in the conserved residues and structural motifs (Table 3.2). It was our expectation that clustering of the mutations would identify regions of the leader RNA important for ligand binding and structure formation in response to lysine; however, no mutational clustering was observed.
### Table 3.2. *lysC* leader RNA variants identified by *in vitro* selection.

Individual isolates from the enriched DNA pool from three rounds of *in vitro* selection were cloned and sequenced. The sequences of the isolates and the mutations identified are shown. The location within the RNA structure and conservation is indicated for each position (Grundy *et al.*, 2003).
Variants with only a single mutation were further analyzed using the RNase H cleavage assay (Fig. 3.5). Each variant was transcribed in the absence and presence of 0.9 mM and 7.5 mM lysine, and RNase H cleavage was used to monitor the formation of helix 1. In comparison to wild type, the variant leader RNAs exhibited RNase H protection patterns that corresponded to three distinct phenotypes, designated Class I, II, and III. Class I variants exhibited a cleavage pattern similar to that of the wild-type RNA, in which helix 1 protection increased as the lysine concentration was increased, with >60% cleavage in the absence of lysine (Fig. 3.5). Class II variants also showed lysine-dependent protection but exhibited partial protection in the absence of lysine. These variants also showed a lysine-dependent increase in protection (Fig. 3.5). Class III variants also exhibited lysine-independent protection of helix 1 (Fig. 3.5). The increased protection of helix 1 observed in the Class II leader RNAs, as compared to wild type, suggests that the mutations in the leader RNA may result in the formation of a structure that more readily forms the terminator structure as compared to wild type. Protection of helix 1 in the absence of ligand, as seen in the Class III variants, suggests that the mutations resulted in changes within the leader RNA structure that cause the RNA to adopt the bound conformation and stabilize helix 1.
Figure 3.5. RNase H cleavage of \textit{lysC} leader variants identified by SELEX.
Secondary structural model of \textit{lysC} ligand binding domain. Arrows denote single nucleotide substitutions identified by SELEX. PAGE purified RNase H cleavage products of SELEX-isolated \textit{lysC} variants transcribed in the presence and absence of lysine. Class I (green) mutants show RNase H cleavage similar to wild type. Class II (purple) mutants are partially protected from cleavage in the absence of lysine. Class III (red) mutants are completely protected from RNase H cleavage at all lysine concentrations.
3.3.6 *In vitro and in vivo* analysis of SELEX variants

To examine the effects of each single nucleotide substitution, representatives of Class II and III variants were transcribed in the presence of lysine using an *in vitro* transcription assay to monitor lysine-dependent termination. It was our expectation that variants with an increased lysine affinity would exhibit termination efficiencies higher than the wild-type construct in the presence of lysine. Transcription of the U63C (Class II) variant resulted in 29% termination in the absence of lysine (Fig. 3.6) which is similar to wild type termination in the absence of lysine (32%). This result was unexpected as the U63C variant exhibited increased protection of helix 1 in the absence of lysine as compared to wild type in the RNase H cleavage assay. We expected that that this variant would more readily form the termination conformation in the absence of lysine and thus be more prone to terminate. However, the ligand-independent termination of the U63C variant is consistent with wild type termination in the absence of lysine. Transcription of the U63C variant in the presence of 0.1 mM and 0.9 mM lysine resulted in 50% and 70% termination, respectively. This result is also consistent with wild type termination in the presence of 0.1 mM (40%) and 0.9 mM (69%). The termination of the U63C variant in the presence of 0.1 mM lysine was similar to wild type and there was no significant increase in lysine-dependent termination. Termination of the U162C (Class III) variant in the absence of lysine was slightly lower than wild type (21%). In the presence of 0.1 mM and 0.9 mM lysine, the U162C variant exhibited 48% and 74% termination, as compared to wild type. The *in vitro* termination results of the variant constructs in the
presence and absence of lysine suggests that the leader RNAs do not have an increased sensitivity for lysine and can terminate transcription similar to wild type when transcribed under the same conditions.

Figure 3.6. *In vitro* transcription of Class II and Class III SELEX variants. The wild type and Class II (U63C) and Class III (U162C) variant representatives with a single base substitution were constructed. Each variant was transcribed in the presence or absence of lysine at the indicated concentrations and lysine-dependent termination was monitored. Percent termination is the amount of terminated transcripts relative to the total transcribed RNA. Reactions were performed in duplicate and reproducibility was ±5%. Error bars represent the standard error of the mean.

To determine if the variant leader RNAs can respond to changes in cellular lysine concentrations, the variants were fused to a *lacZ* reporter gene and integrated into the chromosome of the lysine auxotrophic strain BR151. The wild type and variant construct
from each class was selected and β-galactosidase expression was monitored when cells were grown in the presence and absence of lysine. When cells were grown in the presence of lysine the wild-type construct exhibited repression of lacZ expression whereas when cells were grown in the absence of lysine lacZ expression was induced. The results of the in vivo expression assays of the lysC variants indicated that the mutations caused a loss of lysine-dependent repression (Fig. 3.7). Constitutive lacZ expression was observed in cells harboring the C107A fusion construct. This result was surprising as this Class I representative exhibited RNase H protection similar to wild type and the mutation was in a non-conserved position not predicted to disrupt the regulatory function of the riboswitch. The loss of lysine-dependent repression suggests that the mutation affects the ability of the leader RNA to form the termination conformation in vivo. Although the U63C variant also exhibited a loss of lysine-dependent repression, the basal level of gene expression was higher than wild type but lower than the C107A variant. The increase in expression, as compared to wild type, suggests that the U63C mutation caused a loss of lysine-dependent termination. The reduction in constitutive expression as compared to the C107A variant might be due to the structural conformation that the leader RNA adopts because of the mutation. Similarly, the U162C variant exhibited a loss of lysine-dependent repression in vivo and exhibited expression levels, in the presence and absence of lysine, similar to wild type expression when cells were grown in the absence of lysine (Fig. 3.7). The U162C variant exhibited lysine-dependent termination in vitro similar to wild type and helix 1 protection in the absence of lysine in
the RNase H assay. These results are in contrast to the lack of lysine-dependent repression \textit{in vivo}. Isolates from each class of variants exhibited a loss of lysine-dependent repression \textit{in vivo}. One explanation is that the mutations cause a disruption in the co-transcriptional folding process required for termination. It is interesting that the mutation of non-conserved residues resulted in deregulation of gene expression \textit{in vivo}. In total, this analysis has identified nucleotides that are required for \textit{lysC} regulation but are not conserved. Further analysis will be required to determine how the mutations deregulate gene expression.
Figure 3.7. Lysine-dependent expression of wild type and mutant lysC-lacZ fusions. Transcriptional fusions containing wild type or mutant lysC leader sequences fused to lacZ were introduced into a lysine auxotrophic strain. Cells were grown in the absence and presence of 100 μg/mL lysine and β-galactosidase activity was measured at 1 hour intervals. β-galactosidase activity is expressed in Miller units (Miller, 1972). Reactions were carried out in duplicate and variation was ±15%.

3.3.6 In vitro selection of variant pool using a modified SELEX scheme

The lysC leader variants isolated from the original in vitro selection scheme exhibited in vivo phenotypes that suggested that the mutations that resulted in the protection of helix 1 might be due to an alternate conformation adopted by the RNA. The initial design of the selection scheme included a selective pressure that enriched the
leader RNAs pool based on helix 1 formation. This selective pressure could have potentially enriched a pool of molecules that have structural perturbations that protect helix 1 from RNase H cleavage independent of the lysine concentration. This presented a challenge, as functional riboswitches must perform the ligand-dependent conformational switch that dictates the formation of the regulatory structure. RNAs that do not perform this function are unable to efficiently terminate transcription and regulate gene expression. To enrich a pool of RNAs that can alternate between the two structural conformations in response to lysine, an additional selective pressure was included in a modified SELEX scheme. The additional step required the variant leader RNAs to protect helix 1 in the presence of low lysine concentrations and form the antiterminator in the absence of lysine. This corresponds to the predicted structural rearrangement in which the antiterminator forms in the absence of lysine (Grundy et al., 2003). In the presence of lysine, helix 1 forms which also stabilizes the terminator.

To incorporate this into the SELEX scheme, overlapping DNA oligonucleotides were designed corresponding to positions 209-265 of the leader RNA. PCR using these oligonucleotides extended the DNA template to include the regulatory structures. In the absence of lysine, the leader RNA should form the antitermination conformation in which positions 191-255 form the antiterminator helix. In the presence of lysine, positions 255-265 are expected to form the terminator structure (Fig. 1.10). Therefore, we designed a DNA oligonucleotide complementary to positions 256-265 to probe for antiterminator formation in the absence of lysine. Cleavage of the leader RNA indicates that the
antiterminator has formed, while protection of the leader RNA in the presence of lysine indicates formation of the terminator. To determine if this assay could serve as an additional selective pressure, RNase H cleavage of the wild-type full-length construct was performed. In the absence of lysine, the wild-type construct was cleaved by RNase H and exhibited lysine-dependent protection from cleavage (data not shown). These results demonstrate that the RNase H assay can be modified to monitor formation of the antiterminator. This additional step was incorporated into the in vitro selection scheme and the pool was further enriched (Fig. 3.8).

The variant pool from the previous selection (round 3) was used as the starting template for the dual RNase H selection procedure. After three additional rounds of selection using the modified selection scheme, the isolates from the variant pool were cloned and sequenced. The variants identified contained an average of 2-8 base substitutions per molecule (Table 3.3). Twelve isolates were sequenced and six of the isolates contained mutations in conserved residues (Table 3.3). Most of the mutations were located within the lysine-binding pocket or in the conserved structural motifs found in helix 2. These findings are in contrast to the previous selection that resulted in the identification of variants with only a few mutations in the conserved residues or structural motifs. In addition to the substitution of the conserved residues, mutational mapping of the isolates identified a pattern of mutational clustering. The majority of the isolated variants contained a least one mutation in helix 2 (Table 3.3), and some variants
contained as many as four mutations in helix 2. Most often, these mutations were located in or near the two conserved structural motifs.

The clustering of the mutations suggests that helix 2 is important for the structural transition that is required for lysine-dependent termination. This finding is also consistent with the crystal structure of the L box leader RNA. In the tertiary structure, helix 2 is involved in major tertiary interactions that facilitate the stabilization of the global RNA structure (Garst et al., 2008; Serganov et al., 2008). Mutational analysis of these interactions demonstrated a loss of function when either the kissing interaction or GA (kink-turn) motif was disrupted (Blouin and Lafontaine, 2007). The number of mutations in each variant makes it difficult to predict which mutations contribute to the observed phenotype. Therefore, further biochemical analysis is required to assess the function of each variant and the effects of the mutations.
Figure 3.8. Modified SELEX scheme. Wild-type *lysC* leader DNA was used as a template for mutagenic PCR to create a pool of variant templates. The pool was transcribed in the presence of lysine. A DNA oligonucleotide complementary to the 5’ side of helix 1 was allowed to hybridize and RNAs were cleaved with RNase H. PAGE was used to separate the full-length from the cleaved products. The full-length products that were protected from cleavage were reverse transcribed to make a cDNA copy. The cDNA was amplified by PCR. This DNA was used as the template for transcription and RNase H cleavage of RNAs that form the antiterminator in the absence of lysine. A DNA oligonucleotide that hybridized within the 3’ region of the terminator was used for RNase H cleavage. RNAs that formed the antiterminator were cleaved by RNase H and cDNA was made using reverse transcriptase. The pool was amplified by PCR and the template was used in the next round of selection.
Table 3.3. *lysC* leader variants isolated from additional *in vitro* selection. The enriched DNA pool, from three additional rounds of *in vitro* selection, was cloned and sequenced. Sequences that differ from the wild type *lysC* leader region are indicated. Mutations located in conserved residues are denoted by the degree of conservation as (Grundy et al., 2003).

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Mutation Location</th>
<th>Conservation</th>
<th>Isolate</th>
<th>Mutation Location</th>
<th>Conservation</th>
</tr>
</thead>
<tbody>
<tr>
<td>1_14</td>
<td>U62C helix 2</td>
<td>none</td>
<td>2_28</td>
<td>C60A/G105U helix 2/helix 2</td>
<td>none / none</td>
</tr>
<tr>
<td></td>
<td>U172C binding pocket</td>
<td>&gt;80%</td>
<td>C110G/A112G bdn pkt / bdn pkt</td>
<td>100% / 100%</td>
<td></td>
</tr>
<tr>
<td>1_16</td>
<td>U63C / A67G helix 2 / GA motif</td>
<td>none / &gt;80%</td>
<td>A159G / U164C loop 4 / helix 4</td>
<td>&gt;80% / none</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U83C / G101C GA motif / S-turn</td>
<td>none / &gt;80%</td>
<td>A182G loop 5</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td></td>
<td>C110A / A159G bdn pkt / loop 4</td>
<td>100% / &gt;80%</td>
<td>2_27</td>
<td>A57G / A87C S-turn / GA motif</td>
<td>&gt;80% / none</td>
</tr>
<tr>
<td></td>
<td>A175G helix 5</td>
<td>none</td>
<td>1_17</td>
<td>A71U / U79G loop 2 / loop 2</td>
<td>none / none</td>
</tr>
<tr>
<td></td>
<td>U81A GA motif</td>
<td>none</td>
<td>A95U / A116G S-turn / helix 3</td>
<td>&gt;80% / none</td>
<td></td>
</tr>
<tr>
<td>1_19</td>
<td>A51G / A54G S-turn / S-turn</td>
<td>none / &gt;80%</td>
<td>U62A helix 2</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td></td>
<td>U162C / A179G helix 4 / helix 5</td>
<td>none / none</td>
<td>A71G loop 2</td>
<td>none</td>
<td></td>
</tr>
<tr>
<td>1_21</td>
<td>G66A GA motif</td>
<td>&gt;80%</td>
<td>2_27</td>
<td>A70G / A71G loop 2 / loop 2</td>
<td>none / none</td>
</tr>
<tr>
<td>1_29</td>
<td>U83C GA motif</td>
<td>none</td>
<td>U149C / A177G helix 4 / helix 5</td>
<td>none / none</td>
<td></td>
</tr>
<tr>
<td></td>
<td>G90U helix 2</td>
<td>none</td>
<td>2_27</td>
<td>A57G / A87C S-turn / GA motif</td>
<td>&gt;80% / none</td>
</tr>
</tbody>
</table>

3.4 Discussion

Ligand affinity is a widely studied topic in the riboswitch field. A large number of small molecules are recognized by riboswitches with each class utilizing a different scaffold to recognize and bind the cognate ligand. The investigation of the *lysC* leader RNA affinity is important, as it will broaden our understanding of how the riboswitch
functions within the context of the cell as well as aid in the characterization of a unique lysine-RNA interaction. In this study, we developed an in vitro evolution method, based on SELEX, to select variant leader RNAs with mutations that enhance the binding properties of the lysC leader RNA.

In our preliminary studies, several different methods were utilized to estimate lysine affinity. This analysis differs from the initial estimations that used in-line probing as an indication of affinity. Here our goal was to determine the binding affinity by directly measuring lysine binding to the leader RNA. Unfortunately, size exclusion filtration was not sufficient for this task although it has been successfully used in similar riboswitch analyses (McDaniel et al., 2003; Fuchs et al., 2006; Ontiveros-Palacios et al., 2008). The lysine binding observed when the yitJ leader RNA was included as a negative control suggested that the values obtained were not a result of direct lysine binding but may be due to nonspecific interactions between lysine and the RNA backbone. The increase in radioactivity observed as a function of RNA concentration also supports the notion that lysine may be involved in nonspecific interactions with the RNA. It was predicted that the L box riboswitch has one of the weakest binding affinities in comparison to other riboswitches. Thus, we speculated that the affinity may be too low to be determined using this method. It is also possible that the lysine-RNA complex dissociates so rapidly that formation of the bound complex could not be captured for quantitation. Spectrofluorimetry is an alternative method that we employed to estimate lysine affinity. This method was successfully used in analysis of the SMK riboswitch.
(Smith et al., 2010b). The fluorescent quenching of a site-specifically incorporated 2-AP molecule suggested that 2-AP fluorescence could be monitored as a function of lysine concentration. However, we were unable to detect consistent fluorescent quenching of the 2-AP molecule in response to the increasing lysine concentrations.

Despite the challenges with the affinity determinations, we continued our analysis with an in vitro evolution method designed to identify molecules that form the termination conformation in the presence of reduced lysine concentrations as compared to wild type. This method allowed us to enrich a pool of variant RNAs with a unique function. Mounting biochemical evidence suggests that the formation of helix 1 is one of few structural changes in the lysC leader RNA that indicate the ligand-bound state (Serganov et al., 2008; Blouin et al., 2011). Therefore, we used the RNase H cleavage assay to probe for helix 1 formation in response to lysine. This approach offers several benefits in that it can identify sequences that are critical for both ligand binding and the structural transition. Our initial selection method identified a broad spectrum of single, double, and triple mutants. The majority of the variants contained mutations that were located in non-conserved positions. However, among the twenty-five isolates, eight variants contained mutations in conserved residues. These mutations were located at base of helices 3 and 5, the terminal loops of helices 3 and 4, and the conserved structural motifs. It is interesting however that the mutations in conserved residues were clustered around regions of structural significance with none found in the lysine-binding pocket. The lack of mutations in the lysine-binding pocket may be an indication of the structure-
function relationship between lysine binding and helix 1 formation. It is possible that lysine binding facilitates helix 1 formation and thus mutations that obliterate ligand binding also destabilize helix 1. It is important to note that this effect on helix 1 formation can vary based on the mutation. Our analyses of variants with mutations in conserved residues in the lysine-binding pocket demonstrate that the leader RNA can form helix 1 and terminate transcription despite a loss of lysine recognition (see chapter 4).

Further analysis of the variants by RNase H cleavage revealed three unique classes of variants, Classes I, II and III. The phenotype of each class was based on the RNase H protection as compared to wild type. Class I exhibited protection most similar to wild type while Class II was partially protected from cleavage in the absence of lysine. Class III variants were completely protected from RNase H cleavage as compared to wild type when assayed under the same conditions. Further RNase H analysis supported the initial findings relative to the Class II and Class III variants (data not shown). The ligand-independent protection from cleavage suggested that the mutations might promote stabilization of the termination conformation at lysine concentrations lower than that of wild type. However, lysine-dependent termination of the variants in vitro was similar to wild type with only slight increases in termination. This result was in contrast to our expectations as we anticipated that the Class II and Class III leader RNAs would exhibit a higher level of termination as compared to wild type. In vivo expression analysis revealed that the mutations in the leader RNA caused constitutive expression of the lacZ.
reporter gene in each of the SELEX isolated variants. This is in contrast to the in vitro transcription data and indicates that the leader RNAs are capable of lysine-dependent termination of transcription in vitro. The difference in outcome of the two analysis methods might be directly related to the response of the leader RNAs to the speed of transcription. In our purified in vitro system, the speed of transcription must be reduced, as compared to transcription in vivo, to ensure the recognition of intrinsic pause sites and the proper folding of the leader RNA. In contrast, it was estimated that the speed of transcription in vivo is ~40 nt/sec (Vogel and Jensen, 1994); it is possible that the mutations disrupt a hierarchical folding process required for structure formation. To prevent the enrichment of structurally compromised variants, we modified the in vitro selection method to select for variants that can form the antiterminator in the absence of lysine. This modified selection procedure required that the leader RNAs selected could form the antiterminator in the absence of lysine and the terminator in the presence of lysine. Although some mutations were identified in both selection processes, the majority of the variants isolated using the modified selection scheme harbored mutations that were clustered around the conserved structural motifs in helix 2. This suggests that helix 2 (specifically the conserved structural features) is crucial to the structural transition that is required to bind lysine and regulate transcription.

The initial goal for this work was to identify lysC leader variants with an increased lysine affinity as compared to the wild-type construct. The lack of a reliable assay to measure lysine binding limited our analysis of the lysine affinity. Although we
were unable to reach our original goal, we instead took advantage of the lysine-dependent structural response to monitor the lysine sensitivity of the variant leader RNAs. In this study, we have developed an *in vitro* evolution method for the aptamer domain of the *B. subtilis* lysC leader RNA. This scheme is readily adaptable and can be modified for the analysis of additional riboswitches. This method has identified several non-conserved residues that are required for the lysine-dependent repression of gene expression. Additionally, our data suggests that helix 2 may play an integral role in the ligand-dependent structural transition. Further analysis of the individual variants is required to distinguish between residues that function in ligand binding and recognition as compared to residues that are required for modulation of the RNA structure.
CHAPTER 4

ANALYSIS OF LYSINE RECOGNITION AND SPECIFICITY

4.1 Introduction

Recent work has uncovered a set of bacterial regulatory RNAs, designated riboswitches, in which conserved cis-acting RNA elements located in the “leader region” between the promoter and the regulated coding sequence respond to a variety of physiological signals without additional proteins or co-factors (Henkin, 2008). Typical riboswitches are composed of two domains, the aptamer or ligand-sensing domain, and the expression platform (Smith et al., 2010a). The aptamer binds the cognate ligand and initiates a series of structural rearrangements that affect the fate of the downstream gene through structural modulation of the expression platform. Regulation of gene expression by these conserved leader RNAs can occur by transcription attenuation, modulation of translation initiation, and effects on RNA stability (Cheah et al., 2007; Smith et al., 2010a).

A large number of genes that control essential processes are under riboswitch regulation in Bacillus subtilis (Henkin, 2008). One of the earliest identified riboswitches, designated the L box, regulates transcription of the lysC gene in response to lysine (Grundy et al., 2003; Rodionov et al., 2003; Sudarsan et al., 2003). L box leader RNAs
have been found upstream of a variety of genes involved in lysine biosynthesis, transport, and utilization (Rodionov et al., 2003; Serganov and Patel, 2009a; Jorth and Whiteley, 2010). Lysine biosynthesis in most bacteria, and some plants, occurs through the diaminopimelate pathway. The first step of lysine biosynthesis is the phosphorylation of aspartate by aspartokinase. There are three differentially regulated aspartokinase isozymes in *E. coli* and *B. subtilis*. However, only aspartokinase II, the product of the *lysC* gene, responds to lysine. Formation of aspartate-4-phosphate is critical, as it is also a precursor for methionine and threonine biosynthesis (Patte et al., 1998). Other L box-regulated genes synthesize metabolites in the lysine pathway that are required for other essential physiological processes such as synthesis of the cell wall peptidoglycan and formation of bacterial endospores (Forman and Aronson, 1972; Pavelka and Jacobs, 1996). Additional phylogenetic analyses suggest that L box leader RNAs are also involved in regulation of genes responsible for lysine transport and the regulation of cellular pH.

When lysine concentrations are high, lysine binds to the *B. subtilis lysC* leader RNA and promotes the formation of an intrinsic terminator that prevents transcription of the *lysC* coding region (Grundy et al., 2003; Sudarsan et al., 2003). Low lysine concentrations result in the formation of an alternate antiterminator structure that allows transcription to continue (Fig. 1.12). In contrast, the *E. coli lysC* leader RNA is predicted to regulate expression at the translational level through the formation of a Shine-Dalgarno sequestering helix when lysine concentrations are high (Patte et al., 1998; Grundy et al., 2003).
The lysine analog aminoethylcysteine (AEC) is 10-fold less effective than lysine in promotion of *B. subtilis* lysC termination *in vitro* (Grundy *et al.*, 2003). AEC differs from lysine by a single substitution of carbon with sulfur (Fig. 4.1); the observed differential sensitivity demonstrates specific lysine recognition by the *lysC* leader RNA (Grundy *et al.*, 2003). Lysine analog binding studies using the aptamer region of the *lysC* leader RNA further suggested that leader RNA binding is specific for lysine (Blount *et al.*, 2007).
Figure 4.1. Chemical structures of lysine and closely related lysine analogs. Representations of each lysine analog used in this study are depicted with molecular changes highlighted in yellow. Abbreviated names referred to in the text are in parentheses.
Previous studies of riboswitch specificity identified several unique features adopted by riboswitch RNAs to recognize their cognate ligands. For example, the purine responsive riboswitches use canonical Watson-Crick base pairing for recognition (Noeske et al., 2005). A single mutation of a conserved base in the guanine responsive riboswitch results in a specificity switch that allows recognition of adenine (Barrick et al., 2004). Natural variants of the guanine and adenine riboswitches were identified that selectively bind 2’-deoxyguanosine (dG) and discriminate against guanine (Kim et al., 2007). The discovery of these variants demonstrates that riboswitch RNAs can use similar RNA architectures to recognize different molecules.

Several riboswitches use similar mechanisms to recognize nucleotide-like molecules. In each example, the RNA uses common features such as the face of the nucleotide to assist in recognition. Recognition of SAM by the S box (SAM-I) riboswitch occurs through a base-triple interaction that includes the adenine ring of the SAM molecule (Lu et al., 2010). Similarly, the aminopyrimidine ring of TPP is recognized by the THI-box RNA through molecular stacking with conserved residues (Edwards and Ferre-D'Amare, 2006). The nucleotide face of the cyclic di-GMP (c-di-GMP) molecule is also recognized using asymmetric RNA interactions (Smith et al., 2010c). Although the RNAs recognize similar features of different ligands, each riboswitch utilizes a different mechanism for recognition.
Another common mechanism for ligand recognition incorporates divalent metal ions for both binding and recognition of negatively charged groups. RNAs that have been shown to use this method include the S box (Lu et al., 2010), THI-box (Edwards and Ferre-D'Amare, 2006), glycine riboswitch (Huang et al., 2010), and the glmS ribozyme (Edwards et al., 2007). Analysis of glycine recognition by the Vibrio cholerae riboswitch suggests that Mg$^{2+}$ assisted recognition of the carboxylate group contributes to glycine specificity (Edwards et al., 2007). In addition, glycine analog binding and crystal structure analysis indicate that shape complementarity between glycine and the binding pocket is essential for glycine selectivity. The binding pocket of the leader RNA is precisely structured to only accommodate glycine while the RNA did not bind other small amino acids (Mandal et al., 2004b). Recognition of amino acids by riboswitch RNAs presents an interesting challenge, as the RNA must discriminate against similar molecules found in the cell.

Crystal structure analysis of an L box leader RNA from Thermatoga maritima revealed a complex tertiary structure (Garst et al., 2008; Serganov et al., 2008). The lysine molecule is completely encapsulated by the RNA, and is buried within a lattice formed from conserved core residues. Recent analysis of the lysC leader RNA has highlighted the importance of peripheral elements and long-range interactions in the organization of the lysine-binding pocket; the significance of peripheral elements has also been demonstrated for the S box riboswitch (Blouin et al., 2011: McDaniel et al., 2005). Residues in direct contact with lysine or in the binding pocket are the most highly
conserved. Additionally, a K$^+$ ion was buried between lysine and the RNA (Fig. 4.2) (Serganov et al., 2008). This arrangement is similar to the Mg$^{2+}$ ions found in the glycine-binding pocket of the glycine riboswitch (Edwards et al., 2007). Currently, the role of K$^+$ in the binding and function of L box leader RNAs has not been established, but it was suggested that it might assist in lysine recognition (Serganov et al., 2008).

**Figure 4.2. Tertiary structure of the *Thermatoga maritima* L box lysine-binding pocket.** Lysine binding pocket adapted from Serganov et al. (2008). Nucleotides shown are invariant among all known L box leaders with the exception of G114 (G144 in the *B. subtilis* sequence) which is >80% conserved. Corresponding nucleotides in *B. subtilis* *lysC* are shown in parentheses. Positions mutated for this study are identified with an asterisk. The purple sphere represents a potassium ion.
With the variety of genes under L box control, it is essential that the RNA selectively binds lysine and efficiently regulates expression under the appropriate conditions. The current structural model of the lysine-binding site does not clearly demonstrate how the RNA can specifically recognize lysine and discriminate against substitutions in the ligand. Previous studies of lysine analog recognition focused on a subset of available compounds (Blount et al., 2007). In this study, we test a broad range of lysine analogs, and identify mutations that affect ligand recognition; these mutations reveal molecular interactions responsible for recognition and specificity. We also investigate the role of K\(^+\) in the binding site and the potential effects on leader RNA function.

4.2 Materials and Methods

4.2.1 Construction of DNA templates

Templates for \textit{in vitro} transcription by \textit{B. subtilis} RNA polymerase (RNAP) were generated by fusion of the \textit{B. subtilis} glyQS promoter sequence to the \textit{lysC} leader RNA sequence. PCR was carried out using DNA oligonucleotide primers that included the glyQS promoter and hybridized within the \textit{lysC} leader region. A transcriptional fusion plasmid, pFG328, with a wild type \textit{lysC} insertion (Grundy et al., 2003), was the DNA source for PCR amplification using \textit{Taq} DNA polymerase per the manufacturer’s instructions (Invitrogen, Carlsbad, CA). The resulting PCR fragment contained the \textit{lysC} leader template downstream of the glyQS promoter such that transcription would start at
+17 relative to the *lysC* transcription start-site, and included *lysC* sequence extending to 14 nt downstream of the leader region termination site. The transcription start-site was designed to allow transcription initiation with the dinucleotide ApC and a halt at position G42 (relative to the native *lysC* transcript) during transcription in the absence of CTP. The 3’ end of the construct included an additional 106 bp of random sequence to allow resolution between terminated (252 nt) and readthrough (372 nt) transcripts. PCR products were purified using a QIAquick PCR clean up kit (Qiagen, Chatsworth, CA) and sequenced by Genewiz (South Plainfield, NJ).

DNA templates used for T7 RNAP transcription were generated using complementary pairs of overlapping DNA oligonucleotides as previously described (McDaniel *et al.*, 2005; Yousef *et al.*, 2005). Briefly, the 5’ pair contained the phage T7 RNAP promoter sequence fused to the +18 position of the *lysC* leader region. The remaining complementary pairs contained the wild type *lysC* sequence, each with a 5-nt 3’ overhang complementary to the 5’ region of the adjacent pair. The terminal pair contained the 3’ leader RNA sequence that ended at +234, 3 nt prior to the 5’ position of the transcription terminator. This position was selected so that helix 1 formation could be monitored without the competing antitermination structure. Each internal oligonucleotide pair was phosphorylated using T4 polynucleotide kinase (New England Biolabs, Beverly, MA). The pairs were then mixed, incubated at 95°C, and slow cooled to room temperature for annealing. The paired oligonucleotides were ligated using T4 DNA ligase per the manufacturer’s instructions (New England Biolabs, Beverly, MA). The
resulting DNA template was amplified using the flanking 5’ and 3’ DNA oligonucleotides as primers for PCR using Taq DNA polymerase (Invitrogen, Carlsbad, CA). The final template DNA was purified and sequenced as described above.

4.2.2 Site-directed Mutagenesis

The transcriptional fusion vector pFG328 containing the wild-type lysC DNA template was used as a template for oligonucleotide-directed mutagenesis (Grundy et al., 2003). DNA oligonucleotides containing the desired mutations were designed as primers for PCR amplification of pFG328-lysC DNA using Pfu DNA polymerase (Stratagene, La Jolla, CA). The resulting products were subjected to digestion with DpnI (New England Biolabs, Beverly, MA) to remove the starting wild-type template and introduced into XL-2 blue ultracompetent cells by transformation per the manufacturer’s instructions (Stratagene, La Jolla, CA). The plasmid DNA was isolated using a Promega Wizard prep kit and sequenced to confirm the mutations (Genewiz, South Plainfield, NJ). The final constructs contained the glyQS promoter sequence upstream of the lysC leader RNA with the desired mutations, and were used as the DNA template for in vitro transcription.

The DNA templates for transcription by the T7 phage RNAP contained the wild-type lysC leader RNA sequence downstream of the T7 RNAP promoter sequence. Mutation of the lysC construct was performed by the substitution of complementary oligonucleotide pairs containing the desired mutations. The wild type and mutant pairs were phosphorylated, ligated and amplified by PCR using Taq DNA polymerase
(Invitrogen, Carlsbad, CA). The resulting DNA templates with the appropriate mutations were used in T7 RNAP transcription.

4.2.3 In vitro transcription termination assays

Single round transcription of the wild-type lysC leader RNA template was performed as previously described by Grundy et al. (2002). A reaction mixture of 20 mM Tris-HCl, pH 8, 20 mM NaCl, 10 mM MgCl$_2$, 100 µM EDTA, 150 µM ApC (Sigma), 2.5 µM GTP and ATP, 0.75 µM UTP, [$\alpha$-$^{32}$P]-UTP (GE Healthcare; 0.25 µM, 800 Ci/mmol [30 TBq/mmol]), DNA template (10 nM), and His-tagged purified B. subtilis RNAP (6 nM) was incubated for 15 min at 37°C (Qi and Hulett, 1998). CTP was excluded from the reaction mixture to generate a transcriptional halt at +42 (relative to the native lysC transcription start-site). After incubation, heparin was added to block reinitiation and synchronize transcript synthesis. KCl (27 mM) was added where indicated. Lysine or lysine analogs were added, and elongation was resumed by the addition of 10 µM rNTPs and 40 mM MgCl$_2$ followed by incubation for 15 min at 37°C. The transcription reactions were stopped by phenol/chloroform extraction, and the products were resolved by denaturing PAGE and visualized by PhosphorImager analysis (Molecular Dynamics). Percent termination represents terminated transcripts relative to the total amount of RNA transcribed; this value was plotted as a function of ligand concentration. Non-linear regression analysis was used to determine the concentration
required for half-maximal termination, \((\text{Trm}_{1/2})\) (SigmaPlot, Systat Software). Reactions were performed in duplicate, and reproducibility was ±5%.

### 4.2.4 RNase H cleavage

RNase H cleavage was carried out as described by Grundy et al. (2003). The DNA template for T7 RNAP transcription was generated as described above. Transcription was carried out using a MEGAscriptscript T7 RNAP transcription kit (Ambion, Austin, TX) in the presence of 0.5 μM \([\alpha^{32}\text{P}]\)-UTP (GE Healthcare; 800 Ci/mmol [30 TBq/mmol]) for radiolabelling. RNAs were transcribed in the presence or absence of 7.5 mM lysine or lysine analog, as indicated, for 30 min at 37°C. After transcription, a DNA oligonucleotide (47 μM) complementary to positions 193-204 of the \textit{B. subtilis} lysC leader RNA was added and hybridized for 5 min at 37°C. RNase H (0.45 U; Ambion, Austin, TX) was then added to the reaction and incubated 10 min at 37°C. The reactions were stopped by phenol-chloroform extraction. The resulting RNA products were resolved by denaturing PAGE and visualized by PhosphorImager analysis (Molecular Dynamics). The percentage of RNAs protected from cleavage was calculated as the amount of full length RNAs relative to the total amount of RNA transcribed in each reaction. The reactions were performed in duplicate, and reproducibility was ±10%.
4.3 Results

4.3.1 Wild-type *lysC* transcription termination in response to lysine analogs

*In vitro* transcription of the wild-type *lysC* leader by *B. subtilis* RNAP in the presence of lysine (lys) or lys analogs (Table 4.1) was used to identify features of the lys molecule required for ligand-dependent transcription termination. Lys and lys analogs with the ability to promote termination of the *lysC* leader RNA were further tested *in vitro* using a broad range of ligand concentrations (0-12.8 mM) (Table 4.1).

Transcription of the *lysC* leader region in the presence of lys (10 mM) resulted in 96% termination and the concentration required for half-maximal termination (Trm$_{1/2}$) was 0.17 mM (Table 4.1). *In vitro* transcription in the presence of 11 different lys analogs (added at 10 mM final concentration) showed that alterations to the main chain functional groups, as seen in D-lys and diaminopentane, resulted in a reduction of ligand-dependent termination (Table 4.1). The Trm$_{1/2}$ for both D-lys and diaminopentane, was >13 mM. These results suggest that changes to functional groups that participate in RNA-ligand interactions are detrimental to ligand-dependent termination. Reduced termination in the presence of AEC (84%, Table 4.1) suggests that the sulfur substitution in the side chain of the molecule causes reduced ligand recognition. The Trm$_{1/2}$ in the presence of AEC (0.8 mM) increased 5-fold as compared to lys. The increase in Trm$_{1/2}$ in the presence of AEC as compared to lys demonstrates sensitivity to the presence of the sulfur molecule in AEC. Transcription of the leader RNA in the presence of ligands with more complex
functional groups, as seen in L-arginine (L-arg) and L,L-diaminopimelate (DAP), resulted in a loss of recognition, possibly because the changes result in a molecule too large to fit in the lys binding pocket.

Alteration to the global structure of the molecule, such as the shape of the molecule (hydroxy lys), and the length of the side chain (L-ornithine) caused reduced termination efficiency (Table 4.1). Modification of the ε-amino group was tolerated only in the context of the addition of a methyl group, as in lysine methyl (lys methyl). Transcription in the presence of lys methyl resulted in a Trm_{1/2} of 1.3 mM, an 8-fold increase as compared to lys. The increase in Trm_{1/2} in the presence of lys methyl suggests that the methyl addition to the ε-amino group is deleterious to ligand recognition. The Trm_{1/2} in the presence of lys methyl (1.3 mM) as compared to AEC (0.8 mM) suggests that the sulfur substitution found in AEC is less detrimental to ligand dependent termination than the methyl addition to the ε-amino group (Table 4.1). These results may reflect similarities between the B. subtilis and T. maritima lys binding pockets. In the T. maritima structure, the RNA participated in fewer contacts with the side chain as compared to the ε-amino group (Garst et al., 2008; Serganov et al., 2008). Additionally, analogs with modifications to the carboxyl group resulted in >70% termination. This is consistent with the current structural model of the binding pocket, which indicates the presence of small channels near both the carboxyl and ε-amino groups of lys (Serganov et al., 2008).
The carboxy-modified ligands were more effective than analogs with ε-amino or side chain modifications and exhibited lower \( \text{Trm}_{1/2} \) concentrations (Table 4.1). Transcription in the presence of lysinamide (lysNH), which contains a nitrogen substitution of the hydroxyl carbon, resulted in a \( \text{Trm}_{1/2} \) of 0.3 mM, a 2-fold increase as compared to lys. In the presence of lysNH, the leader RNA exhibited a smaller increase in \( \text{Trm}_{1/2} \) as compared to lys methyl and AEC; this may be indicative of a larger or more stable network of RNA-ligand interactions near the carboxyl group. Substitution of the amide (lysNH) with an ethyl ester (lysEE) caused a 3-fold increase in \( \text{Trm}_{1/2} \) (1.0 mM) as compared to lysNH, and a 6-fold increase relative to lys. The ethyl ester extension of the oxygen may disrupt other essential interactions required for lys-dependent termination. The observed decrease in \( \text{Trm}_{1/2} \) in the presence of lysNH as compared to lysEE may reflect a direct interaction between the leader RNA and the substituted amide. Transcription in the presence of lysine hydroxamate (lysOH) resulted in a 10-fold increase in \( \text{Trm}_{1/2} \) (1.9 mM) relative to lys. Similar to the other carboxy-modified analogs, lysOH contains a hydroxy amide substitution of the hydroxyl carbon (Fig. 4.2). The increase in \( \text{Trm}_{1/2} \) of lysOH as compared to lysNH suggests that the addition of a hydroxyl group causes a reduction in ligand-dependent termination. The observed increase in \( \text{Trm}_{1/2} \) of lysOH as compared to lysEE emphasizes the importance of functional group recognition as compared to minor alterations in ligand length. This indicates that both the identity of the molecular substitution and additional interactions within the binding pocket affect ligand recognition.
Table 4.1. *In vitro* transcription termination of wild-type *lysC* leader in the presence of lys or lys analogs.

<table>
<thead>
<tr>
<th>Lysine modification</th>
<th>Ligand</th>
<th>Wild-type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Term.ᵃ</td>
<td>Ratioᵇ</td>
</tr>
<tr>
<td>Global</td>
<td></td>
<td></td>
</tr>
<tr>
<td>None</td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>Lys</td>
<td>96</td>
<td>3.2</td>
</tr>
<tr>
<td>D-lys</td>
<td>46</td>
<td>1.5</td>
</tr>
<tr>
<td>Diaminopentane</td>
<td>43</td>
<td>1.4</td>
</tr>
<tr>
<td>AEC</td>
<td>84</td>
<td>2.7</td>
</tr>
<tr>
<td>Alkyl chain</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hydroxy lys</td>
<td>35</td>
<td>1.1</td>
</tr>
<tr>
<td>L-ornithine</td>
<td>32</td>
<td>1.0</td>
</tr>
<tr>
<td>Lys methyl</td>
<td>73</td>
<td>2.4</td>
</tr>
<tr>
<td>ε-amino group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>L-arg</td>
<td>35</td>
<td>1.1</td>
</tr>
<tr>
<td>DAP</td>
<td>30</td>
<td>0.97</td>
</tr>
<tr>
<td>Carboxyl group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lysNH</td>
<td>90</td>
<td>2.9</td>
</tr>
<tr>
<td>lysOH</td>
<td>87</td>
<td>2.8</td>
</tr>
<tr>
<td>lysEE</td>
<td>76</td>
<td>2.5</td>
</tr>
</tbody>
</table>

ᵃTermination (term.) is the percent of RNAs terminated in the presence of 10 mM ligand relative to the total amount of RNA transcribed.
ᵇRatio is the ratio of percent terminated transcripts in the presence of 10 mM ligand to the percent terminated transcripts in the absence of ligand. ᶜTrmₜₐₜ is the concentration of ligand required to promote 50% termination. ᵈND, not determined.

4.3.2 Effects of substitutions at G111 and A112

In an effort to identify sequence elements within the *lysC* leader that are important for ligand recognition, we used site-directed mutagenesis to probe RNA requirements for lys-dependent termination. The nucleotides chosen for mutation are highly conserved (Fig. 1.12) and predicted to interact with lys (Fig. 4.1). Universally conserved G111 and
A112 are positioned nearest to the ε-amino group in the binding pocket (Fig. 4.1).

Substitution of G111 with C resulted in reduced termination (12%) in the absence of ligand as compared to wild type (Table 4.1, 31%), and caused a reduction in sensitivity to lys and AEC (Table 4.2). Each G111 variant exhibited a reduction in termination in the presence of lys and AEC. However, the G111A substitution resulted in a 2-fold increase in termination in response to DAP (Table 4.2), which is inactive in the context of the wild-type lysC sequence (Table 4.1). Similarly, the G111U substitution resulted in a 5-fold increase in termination in response to lys methyl, as compared to the no ligand control (Table 4.2). The reduction in lys-dependent termination suggests that G111 plays a pivotal role in lys recognition and/or binding. However, the termination response in the presence of DAP (G111A) and lys methyl (G111U) suggests that G111 is not required for termination. The sensitivity of G111A and G111U to ε-amino-modified analogs is in good agreement with the published crystal structure, as G111 is positioned nearest the ε-amino group of the lys molecule (Serganov et al., 2008).

Unlike G111, in which no substitution allowed normal function, A112 was amenable to substitution with a guanine residue, exhibiting a ligand response similar to that of wild type although termination in the presence of ε-amino and carboxyl modified analogs was reduced (Tables 1 and 2). When A112 was substituted with either pyrimidine base (Table 4.2), there was a decrease in termination in both the presence and absence of ligand (~7% for A112C, and ~19% for A112U). Both A112 and G111 were intolerant to pyrimidine substitutions, which suggests that purines are required for ligand
recognition and lys-dependent termination. In contrast to A112G, the purine-purine substitution of G111 with A was deleterious to the function of the RNA. This suggests that there are specific interactions between G111 and the lys molecule that directly contribute to lys-dependent termination.

Table 4.2. *In vitro* transcription termination of G111 and A112 lysC leader variants in the presence of lys or lys analogs.

<table>
<thead>
<tr>
<th>Lysine modification</th>
<th>Ligand</th>
<th>G111C</th>
<th>G111A</th>
<th>G111U</th>
<th>A112G</th>
<th>A112C</th>
<th>A112U</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Term&lt;sup&gt;a&lt;/sup&gt;</td>
<td>Ratio&lt;sup&gt;b&lt;/sup&gt;</td>
<td>Term</td>
<td>Ratio</td>
<td>Term</td>
<td>Ratio</td>
<td>Term</td>
</tr>
<tr>
<td>Global</td>
<td>None</td>
<td>12</td>
<td>14</td>
<td>10</td>
<td>42</td>
<td>6.8</td>
<td>19</td>
</tr>
<tr>
<td></td>
<td>Lys</td>
<td>13</td>
<td>1.1</td>
<td>14</td>
<td>1.0</td>
<td>9.1</td>
<td>0.89</td>
</tr>
<tr>
<td></td>
<td>D-lys</td>
<td>16</td>
<td>1.3</td>
<td>14</td>
<td>1.0</td>
<td>14</td>
<td>1.4</td>
</tr>
<tr>
<td></td>
<td>Diaminopentane</td>
<td>15</td>
<td>1.2</td>
<td>13</td>
<td>0.92</td>
<td>7.7</td>
<td>0.75</td>
</tr>
<tr>
<td></td>
<td>AEC</td>
<td>12</td>
<td>1.0</td>
<td>13</td>
<td>0.92</td>
<td>10</td>
<td>1.0</td>
</tr>
<tr>
<td>Alkyl chain</td>
<td>Hydroxy lys</td>
<td>13</td>
<td>1.1</td>
<td>12</td>
<td>0.88</td>
<td>11</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>l-ornithine</td>
<td>13</td>
<td>1.1</td>
<td>13</td>
<td>0.92</td>
<td>8.5</td>
<td>0.85</td>
</tr>
<tr>
<td></td>
<td>Lys methyl</td>
<td>13</td>
<td>1.1</td>
<td>13</td>
<td>0.90</td>
<td>51</td>
<td>5.1</td>
</tr>
<tr>
<td></td>
<td>L-arg</td>
<td>14</td>
<td>1.2</td>
<td>13</td>
<td>0.92</td>
<td>11</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>DAP</td>
<td>14</td>
<td>1.2</td>
<td>26</td>
<td>1.9</td>
<td>8.8</td>
<td>0.88</td>
</tr>
<tr>
<td></td>
<td>lysNH</td>
<td>13</td>
<td>1.1</td>
<td>12</td>
<td>0.87</td>
<td>11</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>lysOH</td>
<td>13</td>
<td>1.1</td>
<td>13</td>
<td>0.90</td>
<td>10</td>
<td>1.0</td>
</tr>
<tr>
<td></td>
<td>lysEE</td>
<td>16</td>
<td>1.3</td>
<td>12</td>
<td>0.85</td>
<td>11</td>
<td>1.1</td>
</tr>
</tbody>
</table>

<sup>a</sup>Termination (term.) is the percent of RNAs terminated relative to the total transcription.  
<sup>b</sup>Ratio is the ratio of percent terminated transcripts in the presence of 10 mM ligand to the percent terminated transcripts in the absence of ligand. Reactions were performed in duplicate, and reproducibility was ±5%.
4.3.3 Mutations near the lys carboxyl group result in altered ligand specificity

Lys binding pocket variants were generated to assess the interaction between the leader RNA and the carboxyl region of the lys molecule. Crystal structure analysis revealed base pairing between G40 and C110 of the leader RNA, which is in agreement with previous studies that identified mutations at this position that result in deregulation of the lysC leader and cause AEC resistance (Lu et al., 1991). The conserved G40-C110 and G144-U170 base pairs were substituted with other canonical base pairs and tested for termination in the presence of lys or lys analogs. The G40U-C110A substitution inhibited lysine-dependent termination (Table 4.3). Lys or lys analogs with substitutions in the side chain (AEC) or ε-amino group were not recognized by the G40A-C110U variant (Table 4.3). However, termination of the G40A-C110U variant was observed in the presence of carboxyl-modified analogs. Transcription of the G40A-C110U variant in the presence of lysNH, which replaces a carboxyl oxygen with an amide, resulted in a reduction in termination (Table 4.3) as compared to wild type (Table 4.1). Termination of the G40A-C110U variant increased 4-fold in the presence of lysOH (75%) as compared to the no ligand control (Table 4.3). The Trm½ of the G40A-C110U variant in the presence of lysOH was 2.1 mM as compared to >13 mM for lys (Table 4.4). The Trm½ exhibited by the G40A-C110U variant in the presence of lysOH was similar to wild type; however, the Trm½ was increased, as compared to wild type, in the presence of lys (Table 4.4). The response to lysOH and lack of response to lys suggests that the G40A-
C110U mutations cause a change in ligand recognition specificity. Termination of the G40A-C110U variant in the presence of lysNH as compared to lysOH indicates that the hydroxyl substitution at the carbonyl more effectively promotes termination than an amide at the same position. It is plausible that the change in specificity is due to the formation of new molecular interactions between G40A and the amide or hydroxyl group of the lys analog.
Table 4.3. *In vitro* transcription termination of G40-C110 *lysC* leader variants in the presence of 10 mM lys or lys analogs.

| Lysine modification | Ligand    | Variant 
|---------------------|-----------|---------
|                     |           | U-A     | A-U     |
|                     |           | Term.  | Ratio  | Term. | Ratio |
| Global              | None      | 26     | 17     |
|                     | Lys       | 24     | 0.92   | 17    | 1.0   |
|                     | D-lys     | 34     | 1.3    | 21    | 1.2   |
|                     | Diaminopentane | 25     | 0.96   | 16    | 0.94  |
|                     | AEC       | 22     | 0.85   | 14    | 0.82  |
| Alkyl chain         | Hydroxy lys | 24     | 0.92   | 13    | 0.76  |
|                     | L-ornithine | 22     | 0.85   | 12    | 0.71  |
|                     | Lys methyl | 29     | 1.1    | 21    | 1.2   |
| ε-amino group       | L-Arg     | 34     | 1.3    | 13    | 0.8   |
|                     | DAP       | 26     | 1.0    | 15    | 0.9   |
|                     | lysNH     | 30     | 1.2    | 49    | 2.9   |
| Carboxyl group      | lysOH     | 33     | 1.3    | 75    | 4.4   |
|                     | lysEE     | 24     | 0.92   | 15    | 0.9   |

aTermination (term.) is the percent of RNAs terminated relative to the total transcription.
bRatio is the ratio of percent terminated transcripts in the presence of 10 mM ligand to the percent terminated transcripts in the absence of ligand. Reactions were performed in duplicate, and reproducibility was ±5%.
Table 4.4. Termination efficiency of *lysC* leader variants in the presence of lys or lys analogs.

<table>
<thead>
<tr>
<th>Ligand</th>
<th>Trm&lt;sub&gt;½&lt;/sub&gt; (mM)&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Variant</th>
</tr>
</thead>
<tbody>
<tr>
<td>lys</td>
<td>0.38 ± 0.07</td>
<td>A 112G</td>
</tr>
<tr>
<td>lys</td>
<td>&gt;13</td>
<td>G40A -C 110U</td>
</tr>
<tr>
<td>lys</td>
<td>&gt;13</td>
<td>G144A -U170&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>lys OH</td>
<td>ND</td>
<td>AEC</td>
</tr>
<tr>
<td>lys EE</td>
<td>ND</td>
<td>lysOH</td>
</tr>
<tr>
<td>lys EE</td>
<td>0.85 ± 0.10</td>
<td>lysEE</td>
</tr>
</tbody>
</table>

<sup>a</sup>Trm<sub>½</sub> is the concentration of ligand required to promote 50% termination.

<sup>b</sup>ND, not determined.

The G144-U170 leader variants also exhibited specificity for carboxyl-modified analogs. Substitution of the wild type G-U pair with U-G caused a loss of lys recognition and a reduction in termination in the presence of lysine analogs (Table 4.5). The U170G mutation, which replaces the G144-U170 pair with a G-G pair, resulted in complete loss of ligand-dependent termination *in vitro*, suggesting that base pairing is required at this position (Table 4.5). Variants that maintained base pairing between positions 144 and 170 showed an increase in termination in the presence of lys methyl, lysNH and lysEE. With the exception of the A-U and C-G constructs, all other variants resulted in termination efficiencies lower than that of wild type in the presence of lys methyl, lysNH and lysEE (Tables 1 and 5). The C-G variant (G144C-U170G) showed a 14% increase in termination in the absence of ligand (Table 4.5) as compared to wild type (Table 4.1), which suggests that the nucleotide identity at position 144 affects the unliganded RNA.
structure (C-G). In contrast to the C-G variant, the G-C variant (G144-U170C) resulted in a reduction of ligand-dependent termination as compared to wild type in the absence of ligand (G-C) (Tables 1 and 5). Termination of the G-C variant in the presence of lys and lys analogs was reduced as compared to wild type. These results indicate that the U170C mutation does not abolish function of the RNA but perturbs ligand recognition.

Transcription of the G144A-U170 variant in the presence of lysEE resulted in increased termination (94%) (Table 4.5, A-U) as compared to wild type termination in the presence of lysEE (76%) (Table 4.1). The Trm½ of the G144A variant for lysEE was 0.85 mM, similar to that of wild type (Table 4.4). However, Trm½ of the G144A variant in the presence of lys was >13 mM, suggesting a loss in lys recognition (Table 4.4). The G144A substitution, a G-U to A-U base pair exchange, caused an alternate ligand specificity where the variant leader RNA responds to lysEE and discriminates against lys.

Substitution of the G144A-U170 pair with U-A resulted in a reduction of termination efficiency in the presence of the carboxyl-modified analogs; however, ligand specificity was similar to that of the A-U variant. Recognition of lys methyl by the G144-U170 variants is surprising, as these positions were not predicted to reside near the ε-amino group. One explanation is that the G144-U170 base pair is not primarily responsible for ε-amino group recognition. Termination in the presence of lys analogs as opposed to lys and AEC suggests that substitution of the G40-C110 and G144-U170 base pairs resulted in a loss of lys recognition. However, G144-U170 variants retained specificity for ε-amino and carboxyl modified analogs while G40A-C110U exhibited specificity for
analogs with carboxyl modifications. Together these results indicate that specific mutations of the RNA sequence can result in alternate ligand specificity whereby the leader RNA responds to lys analogs while discriminating against lys.

Table 4.5. *In vitro* transcription termination of G144-U170 *lysC* leader variants in the presence of 10 mM lys or lys analogs.

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Global</td>
<td>None</td>
<td>33</td>
<td>22</td>
<td>45</td>
<td>26</td>
<td>22</td>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Lys</td>
<td>33</td>
<td>1.0</td>
<td>21</td>
<td>1.0</td>
<td>43</td>
<td>1.0</td>
<td>29</td>
<td>1.1</td>
<td>22</td>
<td>1.0</td>
<td>19</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>ε-lys</td>
<td>41</td>
<td>1.2</td>
<td>28</td>
<td>1.3</td>
<td>48</td>
<td>1.1</td>
<td>29</td>
<td>1.1</td>
<td>26</td>
<td>1.2</td>
<td>23</td>
<td>1.1</td>
</tr>
<tr>
<td></td>
<td>Diaminopentane</td>
<td>43</td>
<td>1.3</td>
<td>23</td>
<td>1.0</td>
<td>69</td>
<td>1.5</td>
<td>35</td>
<td>1.3</td>
<td>38</td>
<td>1.7</td>
<td>25</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td>AEC</td>
<td>30</td>
<td>0.9</td>
<td>19</td>
<td>0.9</td>
<td>41</td>
<td>0.9</td>
<td>24</td>
<td>0.9</td>
<td>23</td>
<td>1.0</td>
<td>18</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>Hydroxy lys</td>
<td>38</td>
<td>1.2</td>
<td>23</td>
<td>1.0</td>
<td>51</td>
<td>1.1</td>
<td>24</td>
<td>0.9</td>
<td>30</td>
<td>1.4</td>
<td>18</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>L-ornithine</td>
<td>28</td>
<td>0.8</td>
<td>21</td>
<td>1.0</td>
<td>40</td>
<td>0.9</td>
<td>24</td>
<td>0.9</td>
<td>22</td>
<td>1.0</td>
<td>17</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>Lys methyl</td>
<td>50</td>
<td>1.5</td>
<td>24</td>
<td>1.1</td>
<td>92</td>
<td>2.0</td>
<td>70</td>
<td>2.7</td>
<td>68</td>
<td>3.1</td>
<td>66</td>
<td>3.1</td>
</tr>
<tr>
<td></td>
<td>L-Arg</td>
<td>30</td>
<td>0.9</td>
<td>23</td>
<td>1.0</td>
<td>40</td>
<td>0.9</td>
<td>21</td>
<td>0.8</td>
<td>22</td>
<td>1.0</td>
<td>18</td>
<td>0.9</td>
</tr>
<tr>
<td></td>
<td>DAP</td>
<td>32</td>
<td>1.0</td>
<td>21</td>
<td>1.0</td>
<td>43</td>
<td>1.0</td>
<td>25</td>
<td>1.0</td>
<td>23</td>
<td>1.0</td>
<td>16</td>
<td>0.8</td>
</tr>
<tr>
<td></td>
<td>lysNH</td>
<td>49</td>
<td>1.5</td>
<td>24</td>
<td>1.1</td>
<td>87</td>
<td>1.9</td>
<td>57</td>
<td>2.2</td>
<td>73</td>
<td>3.3</td>
<td>63</td>
<td>3.0</td>
</tr>
<tr>
<td></td>
<td>lysOH</td>
<td>47</td>
<td>1.4</td>
<td>20</td>
<td>0.9</td>
<td>67</td>
<td>1.5</td>
<td>34</td>
<td>1.3</td>
<td>41</td>
<td>1.9</td>
<td>36</td>
<td>1.7</td>
</tr>
<tr>
<td></td>
<td>lysEE</td>
<td>72</td>
<td>2.1</td>
<td>24</td>
<td>1.1</td>
<td>91</td>
<td>2.0</td>
<td>49</td>
<td>1.9</td>
<td>94</td>
<td>4.2</td>
<td>72</td>
<td>3.4</td>
</tr>
</tbody>
</table>

*a*Termination (term.) is the percent of RNAs terminated relative to the total transcription.  
b*Ratio is the ratio of percent terminated transcripts in the presence of 10 mM ligand to the percent terminated transcripts in the absence of ligand. Reactions were performed in duplicate, and reproducibility was ±5%.

4.3.4 K⁺ decreases the concentration of lys required for transcription termination

The *B. subtilis lysC* leader RNA does not require K⁺ for lys-dependent termination *in vitro* (Grundy *et al.*, 2003; Sudarsan *et al.*, 2003). However, recent
analyses suggest that $K^+$ is important for lys recognition and affinity (Serganov et al., 2006; Serganov and Patel, 2009a). We therefore tested the effect of addition of $K^+$ to our purified *in vitro* transcription system. In the presence of $K^+$, the wild-type leader RNA showed a 5-fold decrease in Trm$_{1/2}$ for lys (Fig. 4A). Other monovalent (NaCl) and divalent salts (CaCl$_2$) were also assayed to test whether the observed effect was specific to $K^+$, and no observable change in Trm$_{1/2}$ was detected as compared to wild type transcription in the presence of KCl (S. Mitchell, unpublished results). The 5-fold decrease in Trm$_{1/2}$ for lys in the presence of $K^+$ suggests that $K^+$ plays a key role in lys-dependent termination by increasing the affinity of the RNA for lys.

![Figure 4.3](image)

**Figure 4.3. In vitro transcription of wild type and variant lysC leader RNAs in the presence and absence of potassium.** Wild type or variant leader RNAs were transcribed in the presence or absence of 27 mM KCl. Open symbols represent transcription in the absence of KCl, filled symbols represent transcription in the presence KCl. (A) Wild type (circles) and A112G (squares) templates transcribed in the presence of lys. (B) Transcription of the G40A-C110U variant in the presence of lysOH (circles) and the G144A-U170 variant transcribed in the presence of lysEE (squares). Reactions were performed in duplicate and reproducibility was ±5%.
Variant leader RNAs and lys analogs were further tested to determine if the K$^+$ is involved in ligand recognition. The A112G variant, which does not show altered ligand recognition but exhibited a 2-fold decrease in lys-dependent termination relative to the wild type (Table 4.2), exhibited no observable change in lys sensitivity in response to K$^+$ (Fig. 4A). The lack of response to the presence of K$^+$ may be the result of changes in the structure of the binding pocket imposed by the A112G mutation. Based on the position of K$^+$ in the structure, it is plausible that the G40-C110 or G144-U170 base pairs of the *B. subtilis* lysC binding pocket interact directly with K$^+$ (Fig. 2). The G40A-C110U variant, which exhibited increased termination only in the presence of lysOH (Table 4.3), exhibited no change in Trm$_{5/4}$ for lysOH in the presence of K$^+$ (Fig. 4B). Similarly, the G144A variant, which responds to lysEE, showed no reduction in Trm$_{5/4}$ for lysEE in the presence of K$^+$ (Fig. 4B). Transcription of the G144A variant in the presence of lys demonstrated that K$^+$ was unable to enhance lys recognition (S. Mitchell, unpublished results). These results indicate that K$^+$ does not play a role in the specificity of ligand recognition but rather affects affinity of the wild-type leader RNA for lys.

4.3.5 *lysC* leader RNA variants undergo ligand-dependent structural changes

Structural rearrangement of the leader RNA when the regulatory ligand is present is the hallmark of riboswitch regulation. This structural change is crucial as it directly controls gene expression. The current model of the *lysC* leader RNA structural transition indicates that formation of the transcription terminator requires the stabilization of the
anti-antiterminator (AAT; Figure 1.12) through the formation of helix 1 (Grundy et al., 2003). To monitor the structural change of the variant leader RNAs in response to lys analogs, we used RNase H probing of helix 1. Addition of a DNA oligonucleotide complementary to the 3’ side of helix 1 will result in cleavage only when helix 1 is unpaired; stabilization of the helix prevents binding of the oligonucleotide, which results in protection of the RNA from cleavage by RNase H, an endonuclease that is specific for RNA-DNA hybrids.

Figure 4.4. RNase H cleavage of wild type and variant lysC leader RNAs.
Radiolabelled RNAs were transcribed in the presence of 7.5 mM lys, AEC, lysOH and lysEE and incubated with a DNA oligonucleotide complimentary to the 3’ side of helix 1 to allow hybridization followed by RNase H digestion. Helix 1 protection of wild type in the presence of lys and lys analogs is shown in lanes 1-5; lanes 6-10, A112G variant; lanes 11-15, G40A-C110U variant; and lanes 16-20, G144A variant. FL, full length transcripts protected from RNase H cleavage. C, transcripts cleaved by RNase H. Percent protection is amount of full-length RNAs relative to total amount of RNA transcribed. The reactions were performed in duplicate, and reproducibility was ±10%.
In the absence of ligand, 31% of the wild-type RNAs were protected from RNase H cleavage as compared to 89% protection in the presence of lys (Fig. 4.4, lanes 1 and 2). These results indicate that the presence of lys promotes the formation of helix 1. When wild-type RNAs were transcribed in the presence of high concentrations of AEC, 84% protection was observed (Fig. 4.4, lane 3). Similarly, wild-type RNAs exhibited 82% protection in the presence of lysOH and 60% protection in the presence of lysEE (Fig. 4.4, lanes 4 and 5). The reduced protection of wild-type helix 1 in the presence of lys analogs, as compared to lys, is consistent with lower wild type termination \textit{in vitro} in the presence of lys analogs. For the A112G variant, 48% of the transcribed RNAs were protected from cleavage in the no ligand control as compared to 65% in the presence of lys (Fig. 4.4, lanes 6 and 7). A reduction in protection of the A112G variant, as compared to wild type, was observed in the presence of AEC, lysOH, and lysEE (Fig. 4.4, lanes 8-10). This demonstrates that protection of helix 1 in the A112G construct was more effective in the presence of lys than protection in the presence of lys analogs.

RNase H probing of the G40A-C110U variant (Fig. 4.4, lanes 11-15) resulted in 38% protection in the absence of ligand and 51% protection in the presence of lysOH. Protection of the G40A-C110U variant in the presence of lys (40%) and AEC (33%) decreased 2-fold as compared to protection of wild type in the presence of lys or AEC (Fig. 4.4). The lack of helix 1 formation by the G40A-C110U variant is in agreement with \textit{in vitro} termination results that suggest that the G40A-C110U variant is specific for lysOH and discriminates against lys and AEC. Similarly, protection of the G144A
variant increased 2-fold in the presence of lysEE as compared to the no ligand control (Fig. 4.4, lane 20). Protection of the G144A and wild-type constructs in the presence of lysEE was 48% and 60%, respectively. For each lysC leader variant, stabilization of helix 1 was specific for the lys analogs to which they responded in the in vitro transcription assay, and was reduced in the presence of lys. These results confirm that the ligand-dependent termination effects described above correlate with stabilization of helix 1.

4.4 Discussion

Previous studies of riboswitch specificity have identified RNA elements involved in ligand recognition (McDaniel et al., 2005; Serganov et al., 2006; Kim and Breaker, 2008). However, these features are variable and often specific to the RNA and its cognate ligand. Structural analysis of ligand-bound riboswitches has revealed mechanistic themes employed for ligand recognition. Ligand features such as negatively charged groups and nucleotide-like structures are targets for recognition. Shape complementarity, base pairing, nucleotide stacking, and metal ion coordination are known mechanisms of ligand recognition by riboswitches. The molecular structure of lys makes a number of these mechanisms impossible.

Recent analysis of the aptamer region of the T. maritima L box leader RNA has identified the nucleotide composition of the lys binding pocket. Although the nucleotides that directly contact lys were identified, the molecular interactions and their effect on
function are poorly understood. In this study, we use the lys-responsive *B. subtilis* *lysC* leader RNA to analyze recognition determinants and their effects on ligand-dependent structural changes and transcription termination.

Analysis of transcription *in vitro* in the presence of lys analogs correlates with the crystal structure, which suggested that lys recognition relies heavily on shape complementarity between the ligand and the RNA binding pocket (Serganov *et al.*, 2008). This feature seems to be a common characteristic of riboswitch RNAs, and is also observed in the c-di-GMP and TPP riboswitches (Edwards and Ferre-D’Amare, 2006; Smith *et al.*, 2010c). For the *lysC* leader RNA, the effective concentration for each analog is dependent on its structural similarity to lys. The primary determinant for ligand recognition by the *lysC* leader RNA appears to be the ability of the ligand to fit into the lys binding pocket. This suggests that shape recognition is likely to be the first step in discrimination between lys and similar molecules. Analogs that maintain the overall shape of the molecule, with minor molecular substitutions such as those found in AEC and lysNH, were more effective at promoting both the ligand-dependent structural changes and transcription termination. Changes to the chemistry and charge of the molecule at either the carboxyl or ε-amino termini were tolerated to some degree, and groups at these positions could be replaced with other functional groups that make similar interactions. It is possible that molecular extensions such as a methyl group (lys methyl) or ethyl ester group (lysEE) can fit into small channels in the RNA pocket found near the carboxyl and ε-amino groups. Although these analogs fit into the binding pocket, they
are unable to make the necessary molecular interactions required to promote termination at concentrations similar to those required for lys. This suggests that the lys binding pocket has evolved to bind the lys molecule with high affinity and specificity. These results demonstrate that the \textit{B. subtilis lysC} leader RNA is optimized for lys-dependent regulation.

Mutational analysis uncovered features of the RNA that are essential for lys recognition. G111 is critical for lys-dependent regulation. Previous studies have indicated that this residue is involved in structurally significant base triple interactions that may be essential for formation of the lys binding pocket (Serganov \textit{et al.}, 2008). The data presented here provide evidence that G111 is required for formation of the lys binding pocket as well as lys recognition. It is possible that G111 must be precisely positioned to interact with other nucleotides within the RNA and the bound lys molecule. Additionally, our data suggest that A112 is not essential for ligand recognition but is primarily involved in lys affinity. This effect on affinity may be achieved through extensive contacts with A112 that stabilize the structure of the lys binding pocket.

Variants of the lys binding pocket near the carbonyl region of the lys molecule (G40A-C110U and G144A-U170) exhibited preferential specificity for certain lys analogs and discriminated against lys. The alternate specificity of leader RNA variants suggests that each base pair is responsible for the recognition of molecular features of the lys carboxyl group. Additionally, mutations in the carboxyl region of the binding pocket did not inhibit recognition of the \(\varepsilon\)-amino region, as G144-U170 variants retained the ability to
recognize lys methyl. The leader RNA is able to use alternate molecular interactions with lys analogs to promote termination. However, the increased concentration requirement for lys analogs suggests that the new RNA-analog interactions do not perfectly mimic the interactions with lys. This analysis confirms that the sequence of \textit{lysC} leader RNA is optimized for lys recognition as all of the variants in this study, with the exception of A112G, exhibited a loss of lys-dependent termination. In total, these data indicate that recognition of lys by the L box riboswitch is a modular process in which individual nucleotides recognize independent features of the lys molecule.

The presence of a K$^+$ ion in the L box crystal structure (Serganov \textit{et al.}, 2008) raised questions about the role of this ligand in binding. Analysis of the effects of the addition of K$^+$ to the \textit{in vitro} transcription termination assay revealed that the potassium ion may increase the affinity of the leader RNA for lys. The increased affinity was observed only in the presence of the wild-type RNA and lys. This suggests that both the sequence and structure of the binding pocket are necessary for stimulation of tighter lys binding by K$^+$. Additional studies are required to determine the direct effects of K$^+$ on lys binding.

The discovery of new riboswitches and characterization of RNA-ligand complexes continue to diversify our understanding of the mechanisms used by RNAs to recognize and bind small molecules. Small molecule recognition by riboswitches is essential to regulation, as gene expression must occur only in response to the appropriate
stimulus. Recognition of the appropriate ligand by riboswitch RNAs can occur in a variety of ways with some common themes distributed amongst the different classes. The \textit{lysC} leader RNA utilizes some common mechanisms such as shape complementarity and metal ion assistance for lys binding and recognition, but also makes nucleotide-specific interactions for ligand identification. These interactions differ from the mechanism used by the glycine riboswitch to recognize similar amino acid features (Huang \textit{et al.}, 2010). The alternate specificities exhibited by \textit{lysC} variants demonstrate that the leader RNA uses a unique arrangement of contacts to recognize the ligand and regulate gene expression. This characterization has provided a more detailed view of how the \textit{lysC} leader RNA specifically recognizes the lys molecule and contributes to our current understanding of small molecule recognition by RNAs. However, the specific features responsible for lys binding affinity are still unclear, and require further analysis.
CHAPTER 5

SUMMARY AND FUTURE DIRECTIONS

Riboswitches are highly conserved cis-acting RNA elements that modulate gene expression in response to cellular signals. The regulation of gene expression by riboswitches does not require any additional proteins or other ancillary factors; the only requirements are the RNA and effector molecule (Henkin, 2008). Generally, riboswitches consist of two structural domains, the ligand-binding domain (or aptamer), and the regulatory structure (or expression platform). The ligand-binding domain is a conserved molecular sensor that specifically binds the effector molecule and discriminates against similar ligands. This region is common to all riboswitches and exhibits a significant amount of sequence and structural conservation. Evolutionary conservation of the ligand-binding domain is crucial as it ensures that the RNA folds into a molecular receptor competent for selective binding of the target ligand. Recognition of the effector by the ligand-binding domain results in a structural switch that includes the formation of a regulatory structure. The control of gene expression exerted by the riboswitch is in response to metabolic needs of the cell. Typically, riboswitches are located upstream of genes that are involved in the biosynthesis or recycling of the regulatory effector. If the concentration of the cognate effector is suitable for growth and/or homeostasis, and there is no need for additional synthesis, gene expression is
repressed to prevent the unnecessary expenditure of resources. In contrast, genes involved in the degradation of the effector molecule are activated by the riboswitch in response to the cognate ligand. With this method, the cell has devised an efficient strategy to monitor nutrient availability and modulate gene expression.

The term riboswitch was selected based on the structural change that occurs within the RNA once the effector molecule is bound. Riboswitch folding is critical to regulation because it dictates the formation of a regulatory structure. Although recognition of the effector is known to determine which regulatory structure is formed, the changes in the ligand-binding domain that influence RNA folding remain unclear. Riboswitches can control gene expression during transcription, translation, or mRNA processing (Smith et al., 2010a). In bacteria, premature transcription termination, or inhibition of translation initiation are the most common mechanisms of regulatory control. Regulation that occurs during transcription is the result of the formation of an intrinsic terminator. Formation of the terminator causes RNAP to abort transcription prior to reaching the coding region. Regulation during translation initiation is the result of alternative folding of the riboswitch that prevents translation of the mRNA through sequestration of the RBS. Finally, alternative folding can affect mRNA processing in which the recognition of the effector can occlude or expose an mRNA splice site through modulation of the RNA structure.
Riboswitches are classified based on the effector molecule or regulatory signal to which they respond. Overall, these *cis*-acting RNAs can respond to a variety of molecules such as purines, vitamins, tRNAs, and amino acids (Smith *et al.*, 2010a). The lysine-responsive L box leader RNA was the first of three amino acid-binding riboswitches discovered (Grundy *et al.*, 2003). The discovery of the L box leader RNA revealed one of the largest and most complex riboswitches known to date. At the onset of this project, the L box riboswitch had recently been identified, but the specific sequences and structural features of the RNA remained largely uncharacterized. We focused our analysis on the *B. subtilis* *lysC* leader RNA and approached the project with three main goals. The first goal was to identify the sequence and structural features of the *lysC* leader RNA that are crucial to the lysine-dependent structural transition. This structural transition of the leader RNA is initiated by the binding and recognition of lysine. Our second aim was to elucidate the components of the *lysC* leader RNA that are responsible for the lysine affinity as lysine binding is crucial to the function of the riboswitch. It is known that riboswitches can bind the target ligand with exquisite specificity and discriminate against similar molecules, therefore; our third goal was to identify the features of the *lysC* leader RNA that are responsible for the lysine specificity.

The model of the L box riboswitch structural transition suggests that in the absence of lysine, the leader RNA adopts a structure that prevents stabilization of the ligand-binding domain and terminator through the formation of an antiterminator element. In the presence of lysine, the leader RNA forms an alternate structure that
promotes terminator formation through the stabilization of a central helix known as helix 1 (Grundy et al., 2003). Additionally, X-ray crystallography of the ligand-binding domain revealed a long-range tertiary interaction between helices 2 and 4 in the presence of lysine. In our analysis of the lysine-dependent structural change, we mutagenized conserved sequences in helix 1 and the terminal loop of helix 4 to examine the structural changes known to occur in the lysine-dependent structural transition. Phylogenetic analysis of the L box riboswitch indicates that the length and base pair composition of helix 1 is variable; however, there are two conserved base pairs at the base of the helix near the single-stranded core. Site-directed mutagenesis of the conserved base pairs indicates that the sequence at these positions is crucial to the stability of the leader RNA. Our data suggest that the conserved sequence in helix 1 is crucial to the structural stability required for the alternative folding of the terminator and antiterminator structures. These findings illustrate a direct relationship between the sequence of helix 1 and the regulatory function of the riboswitch. Further analysis of helix 1 is required to determine how the sequences are involved in the folding of the leader RNA. It is possible that the increased termination observed in the helix 1 variants, in the absence of lysine, is a result of increased stabilization of the termination conformation. It is not clear, however, if this increase in termination is a result of destabilization of the antiterminator, or if the additional C-G base pair promotes increased helix 1 stability. Our data suggest that changes in the base pair composition do not significantly contribute to the stability of helix 1, as variants with a G-C pair at the same position did not exhibit
an increase in termination. One approach to determine if the G-U pair destabilizes the antiterminator is to introduce an additional mutation that results in a G-C pair at the same position. If the G-U pair causes destabilization of the antiterminator, the additional mutation is expected to restore the balance between the competing structures by increasing the stability of the antiterminator and restore antitermination in the absence of lysine.

In contrast to the helix 1 variants, loop 4 variants did not disrupt riboswitch function; instead, these variants exhibited reduced transcription termination efficiencies as compared to the wild type construct. This suggests that interactions that are distal to the lysine-binding pocket influence the regulatory function of the riboswitch. This relationship has also been observed in the S box and S\textsubscript{MK} box riboswitches (Haller \textit{et al.}, 2011; Lu \textit{et al.}, 2011). Analysis of the S\textsubscript{MK} box riboswitch indicates that a single mutation in a hypervariable region of the molecule, outside of the binding core, results in the stabilization of an RNA conformation that causes repression of gene expression in the absence of the molecular effector (Lu \textit{et al.}, 2011).

One of the first steps of the riboswitch regulatory pathway is the co-transcriptional process of ligand binding. The physiological role of the riboswitch relies heavily on the binding affinity since the RNA must modulate gene expression in response to specific conditions as indicated by cellular ligand concentrations. The selective binding of the ligand initiates a conformational change within the RNA that ultimately
leads to regulation of gene expression. It is known that the aptamer domain of the various riboswitch classes specifically binds the cognate effector. For most riboswitches, the estimated $K_D$ of the RNA-ligand complex is significantly lower than the intracellular concentration of the effector. Representatives of the TPP riboswitch, estimated to bind to TPP with a picomolar $K_D$, illustrate that riboswitches can detect extremely low intracellular ligand concentrations (Breaker, 2012). In *E. coli*, low nanomolar concentrations of TPP correspond to a cell that may contain only a single TPP molecule.

In comparison to other riboswitch classes, the L box riboswitch has a low affinity for lysine. Therefore, we focused our analysis on the features that are responsible for the ligand-binding affinity. Although we were unable to identify the specific sequences or structures that are directly involved in the observed lysine affinity, we used SELEX to identify sequences and structural features that contribute to ligand binding and the structural transition. Our analysis resulted in the identification of non-conserved nucleotides outside of the binding-pocket that are important for lysine-dependent regulation. Analysis of variants identified by SELEX illustrates that these RNAs were unable to control gene expression *in vivo* despite the lysine-dependent termination exhibited *in vitro*. Additional analysis using a modified SELEX scheme showed that helix 2 is also crucial for the structural transition required for regulation of gene expression. Although the sequences identified by SELEX were non-conserved, it is possible that disruption of helix 2 causes the loss of regulatory control. It is possible that the mutations cause a disruption in a hierarchical folding process required for regulation.
of gene expression. To test this, *in vitro* transcription can be used to monitor RNAP pausing during transcription of the SELEX variants. If the variants are unable to fold properly, it is expected that the transcriptional pausing pattern will deviate from that of the wild-type construct and suggest that the mutations disrupt the co-transcriptional folding of the riboswitch. Other methods such as oligonucleotide-directed RNase H cleavage can be used to monitor structure formation in the presence of lysine. The alternate cleavage patterns of the variant leader RNAs, as compared to the wild-type construct, might be an indication that the mutations cause a disruption in the RNA structure that is detrimental to the lysine-dependent response. Analyses of this type will be instrumental in determining if the mutant phenotype is a result of changes to the sequence, or an effect of the disruption of the RNA structure.

The initial analysis of the L box riboswitch revealed that the leader RNA was specific for lysine and discriminated against the lysine analog AEC (Grundy *et al.*, 2003). The results from this work influenced our analysis of the features of the lysine-binding pocket that are responsible for the lysine specificity. Mutational analysis of conserved sequences and structures of the *lysC* leader RNA resulted in deregulation of gene expression (Grundy *et al.*, 2003). Therefore, we focused our analysis on identifying the conserved sequences and structural features that are crucial to the regulatory function of the riboswitch. The selective binding of the cognate-ligand is just as important to riboswitch function as the ligand affinity. The RNA must be specific to ensure that regulation does not occur in response to the wrong ligand, as this can be detrimental to
survival. Here, we used site-directed mutagenesis to analyze the features of the lysine-binding pocket that are responsible for the lysine specificity. The results of our analysis demonstrate that the lysine-binding pocket is composed of conserved residues that are responsible for recognition of specific features of the lysine molecule. We predict that the leader RNA utilizes a modular recognition approach in which each nucleotide works to ensure that the ligand contains all of the appropriate structural features. Mutation of conserved residues illustrates that the nucleotides in the binding-pocket, with the exception of A112G, are required for lysine-dependent termination. Transcription of the binding pocket variants, in the presence of lysine analogs, also revealed that certain mutations promote an alternate specificity that results in the recognition of lysine analogs instead of lysine.

Crystal structure analysis of the ligand-binding domain revealed a single $K^+$ ion in the lysine-binding pocket (Serganov et al., 2008). It is not clear if $K^+$ is involved in lysine binding or recognition. *In vitro* transcription of wild type and leader RNA variants specific for lysine analogs suggest that $K^+$ does not play a major role in ligand recognition but instead causes a decrease in the lysine concentration required for termination. The increased sensitivity to lysine suggests that $K^+$ might be involved in additional electrostatic interactions with lysine that result in an increased lysine affinity. To determine if $K^+$ increases the ligand binding affinity, isothermal calorimetry is one method that can be used to determine the association and dissociation constants for the RNA-lysine complex in the presence and absence of $K^+$. This will further establish the
role of the K$^+$ ion in the binding pocket and provide a direct measure of lysine binding by the leader RNA.

We have used a variety of genetic and biochemical techniques to characterize the *B. subtilis* *lysC* leader RNA through the identification of sequences and structures involved in lysine binding, ligand specificity, and the structural transition. Specifically, we show that lysine binding and affinity is a sum of the riboswitch parts, and that structural features distal to the ligand-binding site play a role in the structural transition and possibly binding affinity. Our mutational analysis suggests that the lysine-binding pocket is organized so that each nucleotide must recognize a specific molecular feature of the lysine molecule. Mutation of structural features involved in the lysine-dependent structural transition reveals that the sequence of helix 1 is crucial to the alternative folding required for regulation of gene expression. In total, this study highlights the significance of both the evolutionarily conserved and non-conserved elements that are essential to riboswitch function. Continued study of the L box riboswitch from *B. subtilis* and other organisms will be instrumental to our understanding of the minute details and mechanisms that underpin riboswitch regulation.
REFERENCES


Harada, F., and Nishimura, S. 1972. Possible anticodon sequences of tRNA$^{\text{His}}$, tRNA$^{\text{Asm}}$, and tRNA$^{\text{Asp}}$ from *Escherichia coli* B. universal presence of nucleoside Q in the first position of the anticondons of these transfer ribonucleic acids. *Biochemistry* **11**, 301-308.


