Structural and Mechanistic Studies of the THI Box and S\textsubscript{MK} Box Riboswitches

Dissertation

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

By

Angela Mae Smith, M.S.

Graduate Program in Microbiology

The Ohio State University

2009

Dissertation Committee:
Professor Tina Henkin, Advisor
Professor Kurt Fredrick
Professor Michael Ibba
Professor Ross Dalbey
ABSTRACT

Organisms have evolved a variety of mechanisms for regulating gene expression. Expression of individual genes is carefully modulated during different stages of cell development and in response to changing environmental conditions. A number of regulatory mechanisms involve structural elements within messenger RNAs (mRNAs) that, in response to an environmental signal, undergo a conformational change that affects expression of a gene encoded on that mRNA. RNA elements of this type that operate independently of proteins or translating ribosomes are termed riboswitches. In this work, the THI box and SMK box riboswitches were investigated in order to gain insight into the structural basis for ligand recognition and the mechanism of regulation employed by each of these RNAs. Both riboswitches are predicted to regulate at the level of translation initiation using a mechanism in which the Shine-Dalgarno (SD) sequence is occluded in response to ligand binding.

For the THI box riboswitch, the studies presented here demonstrated that 30S ribosomal subunit binding at the SD region decreases in the presence of thiamin pyrophosphate (TPP). Mutation of conserved residues in the ligand binding domain resulted in loss of TPP-dependent repression in vivo. Based on these experiments two classes of mutant phenotypes were identified. Class I mutations resulted in increased accessibility of the SD region and binding of 30S ribosomal subunits regardless of the
presence or absence of TPP. In contrast, Class II mutations resulted in constitutive occlusion of the SD region even in the absence of ligand. The latter class represents the first example of riboswitch mutations that result in stabilization of the ligand-bound conformation when no ligand is present.

For the SMK box riboswitch, mutational analysis verified the importance of conserved residues for binding to S-adenosylmethionine (SAM). A minimal SMK box element that retains the ability to bind SAM was used to determine the high resolution structure of the riboswitch in complex with ligand using x-ray crystallography. The SD sequence was shown to be sequestered in the ligand-bound conformation and is essential for SAM binding, providing one of the first examples of a riboswitch in which the regulatory domain is intrinsic to the ligand binding domain. Specific binding to SAM is dictated through direct and indirect contacts with the adenosine and sulfur moieties of the ligand. These results were verified using a fluorescence assay with a variety of SAM analogs. The half-life of the SAM-RNA complex was shown to be much shorter than the half-life of the mRNA inside the cell, suggesting that the RNA may undergo multiple regulatory events in its lifetime. Finally, nuclear magnetic resonance (NMR) was used to investigate the solution structure of the SMK box riboswitch in the presence and absence of ligand. These experiments provided direct evidence of a structural rearrangement of the RNA in response to ligand, consistent with the model and biochemical data.
This work is dedicated to my mother Glenda Gale Smith.
ACKNOWLEDGEMENTS

I wish to express my deepest gratitude to my advisor, Dr. Tina Henkin, for investing in me as a student and for being an outstanding mentor over the past several years. She has provided me with an excellent working environment and opportunities for research that have enabled me to complete a rewarding and fulfilling dissertation project. I would like to thank her for her support and guidance and for all that she has taught me. I am honored to have worked with such an excellent scientific investigator and teacher. I would also like to thank Dr. Frank Grundy for his support and his contributions to my research. On countless occasions he has provided invaluable discussion and advice on experimental design and research challenges. I am grateful for his scientific insight and for his assistance. I would like to thank other past and present members of the Henkin Lab for their helpful discussion and suggestions over the years. Particularly, I would like to thank Dr. Jerneja Tomsic for providing protocols and helping with my initial training in the Henkin lab. Also, special thanks go to Ryan Fuchs for his help and for the work we shared on the $S_{MK}$ box riboswitch.

I would like to extend special thanks to my committee members Dr. Mike Ibba, Dr. Kurt Fredrick, and Dr. Ross Dalbey who provided me with constant guidance throughout my graduate studies. Their support was especially appreciated during my difficult transition to a new laboratory. I would also like to thank my committee for
helpful discussion regarding my research, in particular regarding the steady-state and
time-resolved fluorescence studies. I am also indebted to the Ibba and Fredrick labs for
the use of their FPLC systems.

I would like to thank all of my collaborators. Specifically, I would like to thank
the lab of Dr. Juan Miranda-Rios for the work we shared on the THI box riboswitch. I
would like to thank the lab of Dr. Ailong Ke for the crystallization and structure
determination of the $S_{MK}$ box riboswitch. Also, I would like to thank Ross Wilson and
Dr. Mark Foster for carrying out the NMR experiments with the $S_{MK}$ box riboswitch.

I would like to thank the lab of Dr. Squire Booker for providing SAM analogs,
the lab of Dr. Joe Krzycki for providing potassium ferricyanide, and the lab of Dr. Bob
Tabita for use of their fluorimeter. I would like to extend a special thanks to Bill Swoager
for providing bacterial strains and for his technical assistance with the fluorimeter and
other equipment used in these studies. Also, I would like to thank Dr. Sarah Walker and
Dr. Kurt Fredrick for providing 30S ribosomal subunits and tRNA$^{fmet}$.

Finally, I would like to thank my family and friends for their endless patience
and support. I would especially like to thank Dr. Herve Roy for his personal as well as
professional assistance. He has been an amazing source of information and scientific
advice on all topics ranging from RNA manipulation to methods in fluorimetry and
chemical kinetics. I am truly thankful for his help.
VITA

October 23, 1977..........................Born-Greenville, South Carolina

1999......................................B.S. Biochemistry, Virginia Polytechnic
Institute & State University

2000-2002..............................Research Associate, Battelle Memorial Institute

2002-present...........................Graduate Teaching and Research Associate,
The Ohio State University

PUBLICATIONS


  *N. Ontiveros-Palacios and A.M. Smith contributed equally to this work.

FIELDS OF STUDY

Major Field: Microbiology
TABLE OF CONTENTS

ABSTRACT ........................................................................................................................ ii

DEDICATION ................................................................................................................... iv

ACKNOWLEDGEMENTS ................................................................................................. v

VITA ...................................................................................................................................... vii

LIST OF TABLES ............................................................................................................ xii

LIST OF FIGURES .......................................................................................................... xiii

LIST OF ABBREVIATIONS ............................................................................................. xvi

CHAPTER 1 .......................................................................................................................1

REGULATION OF GENE EXPRESSION BY RIBOSWITCHES ................................... 1

1.1 Distribution of known riboswitches ................................................................ 4

1.2 Riboswitch classes .......................................................................................... 7

1.2.1 RNA thermosensors ................................................................................. 7
1.2.2 T box riboswitch ..................................................................................... 9
1.2.3 Purine riboswitches ............................................................................ 11
1.2.4 Cyclic di-GMP riboswitch .................................................................... 15
1.2.5 Amino acid riboswitches ...................................................................... 16
1.2.5.1 Glycine riboswitch ............................................................................ 16
1.2.5.2 L box riboswitch ............................................................................... 18
1.2.6 glmS ribozyme .................................................................................. 19
1.2.7 Mg$^{2+}$ sensing riboswitch ................................................................. 22
1.2.8 FMN riboswitch .................................................................................... 23
1.2.9 B$_{12}$ riboswitch .................................................................................. 26
1.2.10 THI box riboswitch ............................................................................... 28
1.2.11 SAM riboswitches ................................................................................. 32
1.2.11.1 S box riboswitch .............................................................................. 33
1.2.11.2 SAM-II riboswitch ........................................................................... 37
1.2.11.3 SMK box riboswitch .......................................................................... 37
1.2.11.4 SAH riboswitch .............................................................................................................. 40
1.3 Project Goals .......................................................................................................................... 41

CHAPTER 2 ..................................................................................................................................... 45
MECHANISTIC STUDIES OF THE THI BOX RIBOSWITCH ........................................ 45
2.1 Introduction ............................................................................................................................... 45
2.2 Materials and Methods ........................................................................................................... 51
  2.2.1 DNA constructs .................................................................................................................. 51
  2.2.2 RNase H probing of the SD region of E. coli thiM transcripts ....... 52
  2.2.3 Toeprint assay of E. coli thiM transcripts with 30S ribosomal subunits .. ........................................... 53
  2.2.4 TPP binding assays .......................................................................................................... 54
2.3 Results ...................................................................................................................................... 55
  2.3.1 In vivo expression analysis of wild-type and mutant THI box RNAs. . 55
  2.3.2 TPP decreases sensitivity of the E. coli thiM SD region to oligonucleotide-directed RNase H cleavage ................................................................. 57
  2.3.3 30S ribosomal subunit binding to the E. coli thiM transcript ............... 60
  2.3.4 Binding of TPP by the E. coli thiM transcript ...................................... 65
2.4 Discussion ............................................................................................................................... 67

CHAPTER 3 ..................................................................................................................................... 74
CRYSTAL STRUCTURE OF THE SMK BOX RIBOSWITCH IN COMPLEX WITH LIGAND ....................................................................................................................... 74
3.1 Introduction ............................................................................................................................... 74
3.2 Materials and Methods ........................................................................................................... 79
  3.2.1 RNA synthesis and purification ....................................................................................... 79
  3.2.2 In vitro SAM binding assays ............................................................................................ 81
  3.2.3 RNase H cleavage assays ............................................................................................... 82
  3.2.4 Apparent $K_D$ determination ......................................................................................... 82
3.3 Results ...................................................................................................................................... 83
  3.3.1 Design and testing of minimal SMK box constructs .................................................... 83
3.3.2 RNase H cleavage analysis of *E. faecalis* met*K* transcripts.................. 85
3.3.3 SAM binding affinity of the SM*K* box riboswitch and a 51-nt minimal
    construct................................................................................................ 87
3.3.4 Crystal structure of the SM*K* box riboswitch ................................. 89
3.3.5 Mutational analysis of residues important for SAM binding............. 94
3.4 Discussion.............................................................................................. 98

CHAPTER 4 ............................................................................................................... 105
INVESTIGATION OF ALTERNATE CONFORMATIONS OF THE SM*K* BOX
RIBOSWITCH USING NMR SPECTROSCOPY................................................. 105
4.1 Introduction............................................................................................. 105
4.2 Materials and Methods.......................................................................... 107
    4.2.1 RNA synthesis and purification.................................................... 107
    4.2.2 In vitro SAM binding assays...................................................... 110
4.3 Results..................................................................................................... 111
    4.3.1 Synthesis and purification of SMK51 RNA ............................... 111
    4.3.2 1D imino proton spectra of SMK51............................................. 114
    4.3.3 2D NOESY spectra of SMK51 RNA in the presence and absence of
        SAM..................................................................................................... 116
    4.3.4 Binding of SAM to the SMK59 transcript................................... 119
    4.3.5 2D NOESY spectra of SMK59 RNA in the presence and absence of
        SAM..................................................................................................... 120
    4.3.6 Resonance assignment using HSQC NMR spectroscopy............... 122
4.4 Discussion............................................................................................... 125

CHAPTER 5 ............................................................................................................... 129
MEASUREMENT OF LIGAND BINDING TO THE SM*K* BOX RIBOSWITCH
USING FLUORESCENCE SPECTROSCOPY..................................................... 129
5.1 Introduction............................................................................................. 129
5.2 Materials and Methods.......................................................................... 131
    5.2.1 RNA constructs............................................................................. 131
    5.2.2 Fluorescence spectroscopy with SAM and related compounds...... 132
5.2.3 Competition binding assays with SAM analogs ........................................... 133
5.2.4 In vivo translational fusions and β-galactosidase assays........................ 133
5.2.5 Measurement of SAM-SMK box complex half-life in vitro...................... 135
5.2.6 Conformational switching of the SMK box riboswitch using fluorescence spectroscopy ........................................................................................................... 135

5.3 Results ............................................................................................................ 136

5.3.1 Design of a bipartite RNA containing a 2-AP substitution.................. 136
5.3.2 Determination of SAM binding using spectrofluorimetry ...................... 138
5.3.3 Analog binding to the SMK box RNA ......................................................... 142
5.3.4 Inhibition of SAM binding by related compounds using size exclusion filtration ................................................................................................................. 146
5.3.5 Repression of SMK box-regulated gene expression by SAM analogs in vivo ................................................................................................................................. 147
5.3.6 Measurement of the stability of the SAM-SMK box complex in vitro 155
5.3.7 Reversibility of the SMK box riboswitch in vitro .................................... 157

5.4 Discussion ........................................................................................................ 159

CHAPTER 6 ................................................................................................................... 163
SUMMARY AND FUTURE DIRECTIONS .................................................................. 163
REFERENCES ............................................................................................................. 170
LIST OF TABLES

Table

1.1 Known classes of riboswitches..........................................................................................4
2.1 Oligonucleotide primers for the *E. coli* thiM leader..................................................52
3.1 Oligonucleotide primers for the *E. faecalis* metK leader...........................................80
4.1 Oligonucleotide primers used for the *E. faecalis* metK construct SMK59...............108
LIST OF FIGURES

Figure

1.1. RNA thermosensors ..................................................................................................... 8
1.2. T box mechanism of gene regulation ................................................................. 10
1.3. Crystal structures of the ligand binding pockets of the guanine and adenine riboswitches ............................................................................................................... 13
1.4. Tandem architecture of the glycine riboswitch in the 5’UTR of the gcvT gene in B. subtilis .......................................................... 17
1.5 The GlcN6P responsive ribozyme in the glmS gene in B. subtilis ......................... 21
1.6. Tandem S box and B12 riboswitches in the 5’UTR of the metE gene in Bacillus clausii ......................................................................................................................... 28
1.7. THI box riboswitch in the 5’UTR of the thiM gene in E. coli ................................. 32
1.8. S box mechanism of transcriptional control ...................................................... 34
1.9. Model of the E. faecalis SMK box mechanism ................................................. 39
2.1. Chemical structure of TPP and secondary structure of the THI box element in the 5’UTR of the thiM gene in E. coli .............................................................. 47
2.2. Structure of the E. coli THI box ligand binding domain in complex with TPP ...... 49
2.3. In vivo expression analysis of wild-type and mutant THI box RNAs ................... 56
2.4. Oligonucleotide-directed RNase H cleavage of E. coli thiM RNA ....................... 59
2.5. Primer extension inhibition analysis of E. coli thiM RNA ...................................... 62
2.6. TPP binding to E. coli thiM transcripts .................................................................. 66
2.7. Proximity of selected residues in the THI box ligand binding pocket .................... 69
3.1. Chemical structures of SAM and SAH and predicted secondary structure of the *E. faecalis* S<sub>MK</sub> box riboswitch.................................................................................................................................................. 76

3.2. Design of a minimal S<sub>MK</sub> box element........................................................................................................................................ 84

3.3. RNase H cleavage analysis of full-length and minimal S<sub>MK</sub> box transcripts................................. 86

3.4. SAM binding affinity of the S<sub>MK</sub> box riboswitch determined by size exclusion filtration.................................................................................................................................................. 88

3.5. Structure of the *E. faecalis* S<sub>MK</sub> box riboswitch in complex with SAM........................... 90

3.6. Key interactions in the S<sub>MK</sub> box ligand binding pocket.................................................................................................................. 92

3.7. Mutational analysis of pairing residues in the P2 and P4 helices......................................................... 95

3.8. Mutational analysis of SAM binding site residues .................................................................................. 97

3.9. Revised secondary structure of the *E. faecalis* S<sub>MK</sub> box riboswitch.............................. 99

3.10. Single domain architecture of the S<sub>MK</sub> box riboswitch..................................................... 103

4.1. Predicted secondary structure of SMK51 and SMK59 used for NMR spectroscopy ........................................................................................................................................................................ 112

4.2. Large-scale purification of S<sub>MK</sub> box riboswitch for NMR spectroscopy......................... 113

4.3. 1D imino proton spectra of SMK51.............................................................................................................. 115

4.4. 2D ¹H-¹H NOESY spectra of SMK51 RNA in the presence and absence of SAM 118

4.5. SAM binding to SMK59 using size exclusion filtration.................................................................................. 120

4.6. 2D ¹H-¹H NOESY spectra of SMK59 RNA in the presence and absence of SAM 122

4.7. Assignment of imino proton resonances of SMK59................................................................................... 124

5.1. Bipartite RNA used to investigate ligand binding to the S<sub>MK</sub> box riboswitch................. 137

5.2. SAM binding to the 2-AP labeled *metK* RNA....................................................................................... 140

5.3. SAM analog structures and apparent *K<sub>D</sub>* values determined by fluorescence spectroscopy.................................................................................................................................................. 143
5.4. Competitive inhibition of SAM binding by related compounds ......................... 147
5.5. Repression of lacZ fusions in B. subtilis upon addition of methionine or SAM ..... 149
5.6. Predicted pathways for the biosynthesis of methionine in B. subtilis ................. 151
5.7. Expression of metK-lacZ translational fusion and growth of cells in the presence or absence of SAM and related compounds ................................................................. 152
5.8. Dissociation of the SAM-SMK box riboswitch complex in vitro ....................... 156
5.9. Conformational switching of the SMK box in response to SAM using fluorescence spectroscopy ................................................................. 159
LIST OF ABBREVIATIONS

Å  angstrom
aa  amino acid
AASD  anti-anti-Shine-Dalgarno
AdoCbl  5’deoxy-5’-adenysylcobalamin
aa-tRNA  aminoacyl-tRNA
AEC  aminoethylcysteine
ASD  anti-Shine-Dalgarno
ATP  adenosine triphosphate
bp  base pair
BSA  bovine serum albumin
cDNA  complementary DNA
CTP  cytidine triphosphate
DAP  dianaminopimelate
DNA  deoxyribonucleic acid
DTT  dithiothreitol
fMet  formyl methionine
FMN  flavin mononucleotide
FPLC  fast protein liquid chromatography
<table>
<thead>
<tr>
<th>Acronym</th>
<th>Full Form</th>
</tr>
</thead>
<tbody>
<tr>
<td>GlcN6P</td>
<td>glucosamine-6-phosphate</td>
</tr>
<tr>
<td>GTP</td>
<td>guanosine triphosphate</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>KB</td>
<td>ketobutyrate</td>
</tr>
<tr>
<td>Kd</td>
<td>dissociation constant</td>
</tr>
<tr>
<td>kDa</td>
<td>kilodalton</td>
</tr>
<tr>
<td>KMTB</td>
<td>2-keto-4-methylthiobutyrate</td>
</tr>
<tr>
<td>LysRS</td>
<td>lysyl tRNA synthetase</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger RNA</td>
</tr>
<tr>
<td>MT</td>
<td>methylthio</td>
</tr>
<tr>
<td>MTA</td>
<td>methylthioadenosine</td>
</tr>
<tr>
<td>MTR</td>
<td>methylthioribose</td>
</tr>
<tr>
<td>NAIM</td>
<td>nucleotide analog interference mapping</td>
</tr>
<tr>
<td>ncRNA</td>
<td>noncoding RNA</td>
</tr>
<tr>
<td>NOESY</td>
<td>nuclear Overhauser effect spectroscopy</td>
</tr>
<tr>
<td>nt</td>
<td>nucleotide</td>
</tr>
<tr>
<td>NTP</td>
<td>nucleotide triphosphate</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>PAGE</td>
<td>polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Definition</td>
</tr>
<tr>
<td>--------------</td>
<td>------------</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>preQ₁</td>
<td>7-aminomethyl-7-deazaguanine</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>ROSE</td>
<td>repression of heat-shock gene expression</td>
</tr>
<tr>
<td>SAC</td>
<td>S-adenosylcysteine</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>SRH</td>
<td>S-ribosylhomocysteine</td>
</tr>
<tr>
<td>sRNA</td>
<td>small RNA</td>
</tr>
<tr>
<td>T₀</td>
<td>time zero</td>
</tr>
<tr>
<td>t₁/₂</td>
<td>half-life</td>
</tr>
<tr>
<td>THF</td>
<td>tetrahydrafolate</td>
</tr>
<tr>
<td>TIC</td>
<td>translation initiation complex</td>
</tr>
<tr>
<td>TIR</td>
<td>translation initiation region</td>
</tr>
<tr>
<td>TMP</td>
<td>thiamin monophosphate</td>
</tr>
<tr>
<td>TPP</td>
<td>thiamin pyrophosphate</td>
</tr>
<tr>
<td>tRNA</td>
<td>transfer RNA</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>-------</td>
<td>----------------------------</td>
</tr>
<tr>
<td>uORF</td>
<td>upstream ORF</td>
</tr>
<tr>
<td>UTP</td>
<td>uridine triphosphate</td>
</tr>
<tr>
<td>UTR</td>
<td>untranslated region</td>
</tr>
</tbody>
</table>
CHAPTER 1

REGULATION OF GENE EXPRESSION BY RIBOSWITCHES

Gene expression is controlled by a complex network of regulatory systems that modulate expression of individual genes during different stages of cell development and in response to changing environmental conditions. Cells continually adjust gene expression to ensure that required gene products are available when they are needed, but that valuable cellular resources are not consumed unnecessarily in the production of gene products that are not required. An understanding of the mechanisms governing these processes is essential for the basic understanding of cell physiology. During gene expression, DNA is transcribed into messenger RNA (mRNA), which is then used as the template for protein synthesis. Regulation can occur at any stage during this process. Early studies into regulatory systems focused primarily on mechanisms involving regulatory proteins. In systems of this type, the regulatory protein often modulates expression of the target gene(s) in response to a cellular metabolite or effector molecule. Subsequently, the activity of the regulatory protein is dependent on the cellular concentration of the effector molecule, and on the physiological conditions present inside the cell.
In recent years it has become apparent that noncoding RNAs (ncRNAs) play a larger role in gene regulation than was previously realized. An increasing number of small RNAs (sRNAs) that are not translated into protein, but that function as regulators of gene expression, have been discovered (Storz et al., 2005). Some ncRNAs are at least partially complementary to their target mRNAs and regulate expression in *trans* through base pairing interactions that can affect translation efficiency or message stability. Other ncRNAs interact directly with regulatory proteins or the transcription/translation machinery to affect expression levels (Babitzke and Romeo, 2007; Wassarman, 2007). Regulation conferred by these types of RNAs is dependent on their abundance inside the cell, which is determined by their own expression levels and turnover rates.

Another class of regulatory RNAs reside in the untranslated regions (UTRs) of mRNAs and respond directly to environmental signals to regulate expression of *cis*-encoded genes without the requirement for any additional regulatory factors or proteins (Henkin, 2008). These RNA elements, called riboswitches, typically utilize a mechanism in which signal recognition results in a structural rearrangement of the mRNA that alters the efficiency of expression of the downstream open reading frame (ORF). The two most common mechanisms employed by riboswitch RNAs operate at the level of either premature transcription termination (i.e., transcription attenuation) or translation initiation.

Riboswitches that regulate at the level of premature transcription termination undergo a conformational switch that leads to either stabilization or destabilization of an intrinsic transcription terminator helix in response to ligand binding. When the terminator helix is not stabilized (either in the presence or absence of ligand depending on whether
the riboswitch is an “on” or “off” switch), an alternate helix, called the antiterminator, is formed. Ligand binding to the riboswitch RNA determines which of these two mutually exclusive structures (terminator or antiterminator) is favored.

Riboswitches that operate at the level of translation initiation use a similar mechanism in which ligand binding affects accessibility of the Shine-Dalgarno (SD) sequence (located in the ribosome binding site [RBS]) that is required for initiation of translation of the downstream ORF. In these systems, when gene expression is downregulated, the SD sequence is occluded through base pairing interactions with a complementary upstream sequence, the anti-Shine-Dalgarno (ASD) sequence. When gene expression is on, the ASD sequence pairs with a third sequence, the anti-anti-Shine-Dalgarno (AASD) sequence, and the SD sequence is free to interact with the translation initiation complex (TIC). In addition to these two types of riboswitch-mediated regulatory control, riboswitch elements that utilize mechanisms that affect mRNA stability and transcript splicing have recently been discovered (Bocobza et al., 2007; Cheah et al., 2007; Kubodera et al., 2003; Sudarsan et al., 2003a).

Riboswitch RNAs bind their cognate ligand with extreme selectivity and are finely tuned to respond to ligand concentrations that exist inside the cell. These elements demonstrate the high degree of structural versatility of RNA molecules, despite their relatively simple composition compared to proteins. This chapter will focus on known classes of riboswitch RNAs, mechanisms of regulation employed by these RNAs, and the structural basis for specific ligand recognition.
1.1 Distribution of known riboswitches

Riboswitches are classified by the environmental signal specifically recognized by the RNA as well as the highly conserved sequence and structural elements that usually comprise the region of the molecule involved in signal recognition. RNAs that recognize small metabolites make up the majority of known riboswitch classes. However, T box riboswitches that recognize transfer RNAs (tRNAs) and thermosensor RNAs that respond to changes in temperature have also been identified. A list of known riboswitch classes is shown in Table 1.1.

<table>
<thead>
<tr>
<th>Riboswitch</th>
<th>Molecular signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA thermosensors</td>
<td>change in temperature</td>
</tr>
<tr>
<td>T box</td>
<td>uncharged tRNA</td>
</tr>
<tr>
<td>G box</td>
<td>guanine</td>
</tr>
<tr>
<td>adenine</td>
<td>adenine</td>
</tr>
<tr>
<td>2’-dG</td>
<td>2’-deoxyguanosine</td>
</tr>
<tr>
<td>preQ1</td>
<td>7-aminomethyl-7-deazaguanine</td>
</tr>
<tr>
<td>cyclic di-GMP</td>
<td>cyclic di-GMP</td>
</tr>
<tr>
<td>glycine</td>
<td>glycine</td>
</tr>
<tr>
<td>L box</td>
<td>lysine</td>
</tr>
<tr>
<td>glmS ribozyme</td>
<td>glucosamine-6-phosphate</td>
</tr>
<tr>
<td>M box</td>
<td>magnesium</td>
</tr>
<tr>
<td>FMN</td>
<td>flavin mononucleotide</td>
</tr>
<tr>
<td>B₁₂</td>
<td>5’deoxy-5’-adenosylcobalamin</td>
</tr>
<tr>
<td>THI box</td>
<td>thiamin pyrophosphate</td>
</tr>
<tr>
<td>S box (SAM-I)</td>
<td>S-adenosylmethionine</td>
</tr>
<tr>
<td>SAM-II</td>
<td>S-adenosylmethionine</td>
</tr>
<tr>
<td>SₘK box (SAM-III)</td>
<td>S-adenosylmethionine</td>
</tr>
<tr>
<td>SAH</td>
<td>S-adenosylhomocysteine</td>
</tr>
</tbody>
</table>
Most riboswitch RNAs discovered to date regulate genes involved in metabolism and transport of vitamins, amino acids, nucleic acids, enzyme cofactors, and metal ions (Table 1.1). In many instances, the ligand recognized by an individual riboswitch is a metabolic end-product of the regulated pathway, providing a form of feedback inhibition, such that increased ligand levels promote downregulation of gene expression. In contrast, some riboswitches respond to a substrate or byproduct of the regulated metabolic pathway, such that signal recognition leads to upregulation of gene expression. Conservation of sequence elements important for ligand recognition across diverse species suggests a shared origin for members of individual riboswitch classes. However, in some instances several distinct classes have evolved that utilize different mechanisms for recognition of the same molecular signal, as in the case of the SAM binding riboswitches, suggesting independent evolutionary events (Wang and Breaker, 2008).

While most known riboswitches have been identified in bacteria, one riboswitch, the THI box, has been found in all three domains of life (Miranda-Rios et al., 2001; Sudarsan et al., 2003a). Genome-wide searches for known and novel classes of riboswitch RNAs is complicated by the fact that some conserved elements required for riboswitch function are important for maintaining secondary and tertiary structural features, but can vary in primary sequence. Additionally, conserved structural motifs involved in signal recognition sometimes contain long regions of intervening sequence between key elements (e.g., in the S_{MK} box riboswitch) that make recognition difficult, and can also be located at a distance from the site of the regulatory event. Early discoveries of riboswitch RNAs were accomplished by manual searches for conserved
sequence and structural elements in the 5'UTRs or leader sequences of genes with no known regulatory system. While improvements in bioinformatics have yielded automated methods for identifying putative structured RNAs in noncoding regions of mRNAs, identification of new riboswitches remains problematic due to their structural diversity.

Recent advances in metagenomics provide an opportunity for investigating the prevalence of individual metabolite binding riboswitches in entire microbial populations. A recent metagenomics analysis revealed the relative abundance of known and putative riboswitch classes in genomic libraries obtained from three different microbial communities (Kazanov et al., 2007). The abundance patterns for individual riboswitch RNAs in this study were generally in close agreement with results obtained from analysis of complete bacterial genomes, with some differences noted between the three environmental samples. In general, the THI box, B12, and glycine riboswitches were the most abundant, the SAM responsive and FMN responsive riboswitches were moderately abundant, and the \textit{glmS}, lysine, and purine riboswitches were the least abundant. These trends reflect not only the occurrence of individual riboswitch elements in bacterial genomes, but also the prevalence of certain bacterial species within each microbial community. While these types of studies provide valuable information regarding the relative frequency and distribution of riboswitch classes in ecological systems, a comprehensive survey of existing RNAs using these methodologies is complicated by the same challenges listed above.
1.2 Riboswitch classes

1.2.1 RNA thermosensors

RNA thermosensors are regulatory RNAs that modulate gene expression in response to changes in temperature (Narberhaus et al., 2006). These riboswitch elements are unique in that they do not recognize a chemical signal, but instead directly monitor a physical parameter in the environment. Thermosensor RNAs are often encoded in the 5’UTR of genes involved in heat- and cold-shock response pathways. To date, all known RNA thermosensors regulate gene expression at the level of translation initiation through formation of a helix that affects accessibility of the translation initiation region (TIR) to the ribosome. In most cases, when environmental temperatures are sufficiently low, the TIR is occluded and is not accessible for interaction with the TIC (Fig. 1.1). As environmental temperatures rise, the RNA secondary structure is destabilized thereby exposing the TIR for translation initiation and inducing gene expression. A translational mechanism of regulatory control ensures that the mRNA transcript is available for translation initiation as soon as gene products are required, allowing for a more rapid response. This mechanism is utilized by one of the most common thermosensor RNAs called the ROSE (repression of heat-shock gene expression) element which is predicted to be responsible for the regulation of heat-shock genes in numerous alpha and gamma proteobacteria (Narberhaus et al., 1998; Waldminghaus et al., 2005). Another example of an RNA thermosensor that utilizes this mechanism of regulation is located in the 5’UTR of the \textit{E. coli} rpoH transcript, which encodes the heat-shock sigma factor, $\sigma^{32}$ (Morita et al., 1999a; Morita et al., 1999b). Increased levels of $\sigma^{32}$ result in increased expression of a set of genes that are essential for heat-shock response in \textit{E. coli}. 

7
Fig. 1.1. RNA thermosensors. Thermosensor RNAs regulate gene expression at the level of translation initiation. In most instances, the SD sequence is occluded at low temperatures (left), and translation initiation is inhibited. When temperatures increase, the SD sequence becomes accessible, and gene expression is turned on. Adapted from (Henkin, 2008).

In addition to regulating genes involved in heat- and cold-shock response, thermosensor RNAs regulate expression of virulence factor genes in organisms that encounter variable environmental conditions during the course of host cell infection. One example is the thermosensor in the 5’ UTR of the prfA gene in Listeria monocytogenes. Expression of the transcriptional activator PrfA at 37 °C, but not at 30 °C, results in induction of L. monocytogenes virulence genes when the pathogen has entered into a viable mammalian host (Johansson et al., 2002). Structural probing, mutational analysis, and in vivo translational fusion experiments confirmed that increased temperature induces a conformational change in the prfA leader RNA that results in increased accessibility of the TIR. A similar mechanism is postulated for regulation of the virulence gene transcriptional activator lcrF in Yersinia pestis (Johansson and Cossart, 2003).
While the most common mechanism of regulation by thermosensors involves melting of mRNA secondary structure with increasing temperatures, this is not the only mechanism utilized by these regulatory RNAs. In the cIII gene in bacteriophage lambda the mRNA TIR is accessible at 37 ºC, but becomes occluded in an alternative conformation that is stabilized at 45 ºC (Altuvia et al., 1989). This provides an example of an RNA thermosensor that turns gene expression off in response to increasing temperature and demonstrates that these RNA elements can exploit changes in temperature to perform a variety of functions.

1.2.2 T box riboswitch

T box RNAs are unique riboswitches that interact directly with tRNAs to modulate expression of amino acid-related genes (Grundy and Henkin, 1993; Gutierrez-Preciado et al., 2009). These RNA elements, which appear frequently in Gram-positive bacteria and occasionally in Gram-negative bacteria, usually reside in the 5’UTR of genes encoding aminoacyl-tRNA synthetases and proteins involved in amino acid biosynthesis and transport (Grundy and Henkin, 1993, 2003). T box riboswitches most commonly utilize a mechanism of transcription termination control in which binding of a specific uncharged tRNA leads to induction of gene expression (Fig. 1.2). Typically, in the absence of uncharged tRNA the T box RNA favors a conformation in which an intrinsic transcription terminator is formed, thereby preventing transcription of the downstream coding sequence. Interaction with uncharged tRNA results in disruption of the terminator helix and stabilization of an antiterminator element, leading to increased transcription of the full-length mRNA. T box elements that are predicted to regulate at
the level of translation initiation by modulating accessibility of the SD sequence have been identified in some Actinobacteria (Gutierrez-Preciado et al., 2009; Seliverstov et al., 2005; Vitreschak et al., 2008).

![T box mechanism of gene regulation](image)

**Fig. 1.2. T box mechanism of gene regulation.** The T box riboswitch commonly regulates gene expression at the level of transcription termination in response to the charging state of the cognate tRNA. When the cognate tRNA is charged with an amino acid (aa), it fails to make key contacts with the antiterminator element via its acceptor stem. Under these conditions a terminator helix (T) is favored and transcription is halted (left). When the cognate tRNA is uncharged, it contacts the riboswitch RNA via residues in both its acceptor stem and anticodon stem-loop, thereby stabilizing an antiterminator element (AT) and inducing gene expression (right). Adapted from (Grundy and Henkin, 2004).

While all T box RNAs exhibit common structural features, recognition of uncharged tRNAs is specific for each member of the T box family. For example, the T box element that is located in the 5’UTR of the *Bacillus subtilis tyrS* gene, which encodes tyrosyl-tRNA synthetase, responds only to tRNA^{Tyr} and not to other tRNAs. This
specificity is achieved primarily through base pairing between the anticodon of the tRNA and a complementary sequence in the T box RNA called the Specifier Sequence (Grundy and Henkin, 1993; Grundy et al., 2002). Additional interactions between unpaired residues (NCCA) at the 3’ end of the tRNA acceptor stem and residues (UGGN, where the N residues covary) located in a conserved bulge in the antiterminator of the T box RNA provide additional specificity, and ensure that only uncharged tRNA induces gene expression. While aminoacylated-tRNA (aa-tRNA) is able interact with T box RNA via its anticodon, interaction at the acceptor stem is blocked when an amino acid residue is attached to the 3’ end of the tRNA. Because of this partial interaction, aa-tRNA acts as a competitive inhibitor for binding of uncharged tRNA to the T box riboswitch, and levels of gene expression are determined by the ratio of charged to uncharged tRNA (Grundy et al., 1994; Yousef et al., 2005).

1.2.3 Purine riboswitches

Purine riboswitches have been identified in a variety of bacteria including many Firmicutes and some members of the Fusobacteria, Proteobacteria, and Tenericutes. There are four known classes of riboswitch RNAs that specifically recognize purines or modified purines (Kim and Breaker, 2008). These regulatory RNAs usually reside in genes involved in purine metabolism and transport. Three of the purine-responsive riboswitches exhibit remarkable structural similarity within the ligand binding domain and selectively recognize guanine, adenine, or 2’-deoxyguanosine (2’-dG), while the fourth class uses a different architecture to bind the guanosine analog 7-aminomethyl-7-deazaguanine (preQ₁).
The first purine riboswitch to be discovered was the G box riboswitch, which specifically recognizes guanine (Mandal et al., 2003). Most of the guanine sensing riboswitches operate at the level of transcription termination using a mechanism in which ligand binding represses gene expression. The G box riboswitch that is located in the 5′UTR of the xpt-pbuX operon in B. subtilis was shown to bind guanine with an apparent $K_D$ of ~5 nM. Alteration of any of the guanine functional groups results in a dramatic decrease in binding affinity, since riboswitch recognition involves hydrogen bonding interactions at each of these positions. A canonical Watson-Crick base pair between guanine and a conserved cytidine residue (C74) in the G box ligand binding domain is essential for riboswitch specificity.

Adenine binding riboswitches were initially identified as a subset of the G box riboswitches because of structural similarities between the RNAs. As in the case of the G box system, a Watson-Crick base pair between a conserved residue (U74) in the RNA sequence and the adenine ligand is essential for specificity in the adenine riboswitch. Substitution of the uridine at position 74 by cytidine is sufficient to alter the specificity of an adenine riboswitch to that of a guanine binding riboswitch and vice versa (Mandal and Breaker, 2004). In contrast to the G box system, most adenine riboswitches activate gene expression in response to ligand binding. Some adenine riboswitches, such as the ydhL riboswitch from B. subtilis, operate at the level of transcription antitermination, while others, such as the add riboswitch from Clostridium perfringens, are predicted to affect translation initiation (Mandal and Breaker, 2004; Serganov et al., 2004). The adenine riboswitch that is located in the leader of the B. subtilis ydhL gene binds its ligand with an apparent $K_D$ of ~300 nM. The ydhL gene is predicted to encode a purine efflux pump that
helps maintain the concentration of purines inside the cell. Therefore, when adenine levels are high $ydhl$ expression is upregulated and purines are pumped out of the cell at a higher rate.

Crystallization of the ligand binding domains of the guanine riboswitch located in the 5'UTR of $xpt-pbuX$ operon in $B.~subtilis$ and the adenine riboswitch from the $add$ gene in $Vibrio~vulnificus$ allowed direct comparison between the two ligand-bound structures (Serganov et al., 2004). As predicted, the ligand binding pockets of the two riboswitches are nearly identical and specificity in each case is dictated by the residue (at position 74) that base pairs with the appropriate ligand (Fig. 1.3). Several other conserved residues make direct hydrogen bonds with the ligand in each of these riboswitches.

**Fig. 1.3. Crystal structures of the ligand binding pockets of the guanine and adenine riboswitches.** The ligand binding pockets of the guanine and adenine riboswitches are remarkably similar. The primary difference between the two is the identity of the residue at position 74 in the riboswitch sequence, which is a C in the G box and a U in the adenine riboswitch. This residue makes a canonical Watson-Crick base pair with the ligand in each class. Adapted from (Serganov et al., 2004). Images generated from coordinates deposited in the Protein Data Bank under accession numbers 1Y26 and 1Y27.
The third class of purine-responsive riboswitch, which binds 2’-dG, contains conserved residues in the core of the ligand binding domain that differ from those found in the guanine and adenine riboswitches (Edwards and Batey, 2009; Kim et al., 2007). In addition, the 2’-dG riboswitch exhibits altered tertiary contacts that are predicted to affect folding of the ligand binding domain and allow for accommodation of the deoxyribose group. The 2’-dG-responsive riboswitch has been identified only in a single species, *Mesoplasma florum*, and is located in genes encoding a phosphate transporter and the enzyme ribonucleotide reductase, which is involved in the conversion of ribonucleotides into deoxyribonucleotides (Kim et al., 2007). This riboswitch has been shown to regulate gene expression through ligand-dependent stabilization of an intrinsic terminator helix. No high resolution structure is currently available for the wild-type 2’-dG riboswitch. However, structural studies using a guanine riboswitch containing limited sequence substitutions that alter ligand specificity have provided clues to how 2’-dG is recognized in this system (Edwards and Batey, 2009).

The fourth type of purine binding riboswitch recognizes 7-aminomethyl-7-deazaguanine, also called preQ₁. preQ₁ is a precursor for queuosine, which is a modified nucleoside that is incorporated at the wobble position of certain tRNAs, and is important for maintaining translational fidelity (Meier et al., 1985). The preQ₁ riboswitch, which is relatively small compared to other metabolite binding riboswitches, is located in genes involved in queuosine biosynthesis. The RNA element in the 5’ UTR of the queC gene in *B. subtilis* has been shown to bind preQ₁ with an apparent K_D of ~20 nM (Roth et al., 2007). The recent determination of the high resolution structure of a minimal 34-nt preQ₁ riboswitch element provided insight into the structural features required for selective
ligand recognition (Klein et al., 2009). Although this RNA forms a unique structure relative to the other purine binding riboswitches, it also exploits a canonical Watson-Crick base pair for recognition of its cognate ligand.

1.2.4 Cyclic di-GMP riboswitch

Cyclic di-GMP (c-di-GMP), which is comprised of two covalently linked guanosine monophosphate residues, acts as a second messenger in certain bacterial signal transduction pathways (Hengge, 2009). The cellular concentration of c-di-GMP depends on the activities of two enzymes, diguanylate cyclases (DGC), responsible for the synthesis of cyclic di-GMP, and specific phosphodiesterases (PDEs), responsible for its breakdown. c-di-GMP interacts with a variety of regulatory proteins, and also binds directly to mRNA elements to regulate expression of genes involved in a variety of basic cellular processes. Riboswitch elements that are predicted to specifically recognize c-di-GMP (GEMM elements; genes for the environment, for membranes, and for motility) have been identified in many genomes of eubacterial species including members of the Firmicutes, Proteobacteria, Bacteroidetes, and Planctomycetes (Sudarsan et al., 2008; Weinberg et al., 2007). Bacterial processes that are regulated by the GEMM element include flagella and pili production, biofilm formation, and virulence gene expression. Individual GEMM elements are predicted to be differentially affected by ligand binding, with some inducing gene expression and others repressing expression in response to c-di-GMP. Also, some of these riboswitches are predicted to regulate at the level of translational control, while others have been shown to regulate transcription termination (Sudarsan et al., 2008).
1.2.5 Amino acid riboswitches

1.2.5.1 Glycine riboswitch

Conserved sequence motifs corresponding to a glycine binding riboswitch have been identified in many bacterial species (Barrick et al., 2004; Mandal et al., 2004). These elements frequently reside in the 5’UTRs of genes involved in the degradation of glycine for use as an energy source. In the case of the *B. subtilis gcvT* operon, gene expression is upregulated in response to ligand binding due to destabilization of an intrinsic transcription terminator. This riboswitch has also been identified in the 5’ UTR of the *Vibrio cholerae VC1422* gene, which encodes a putative sodium and alanine transporter.

Glycine riboswitches frequently uses a tandem arrangement of two adjacent ligand binding domains working in concert to respond to subtle changes in the intracellular levels of glycine (Mandal et al., 2004) (Fig. 1.4). While the core residues comprising the ligand binding domain of each of the adjacent RNA elements are highly conserved, there are structural differences between the two, and the motifs are referred to as Type I and Type II. In the *V. cholerae VC1422* motif, ligand binding occurs with a 1:1 stoichiometry at each RNA element, and binding is cooperative such that binding at one site increases affinity for glycine at the other site. Nucleotide analog interference mapping (NAIM) revealed regions in the RNA that are important for maintaining tertiary interactions and cooperativity between the two domains of a glycine riboswitch from *Fusobacterium nucleatum* (Kwon and Strobel, 2008). Additionally, small-angle X-ray scattering provided low resolution structural models of the glycine riboswitch in different Mg$^{2+}$ conditions, and in the presence and absence of ligand (Lipfert et al., 2007). The
glycine riboswitch represents the only known example of a riboswitch that couples cooperative ligand binding at two sites to a single regulatory response. In other instances where riboswitches occur in tandem, each ligand binding domain is coupled to its own regulatory output domain, such that each element acts independently.

Fig. 1.4. Tandem architecture of the glycine riboswitch in the 5’UTR of the \textit{gcvT} gene in \textit{B. subtilis}. The glycine riboswitch commonly assumes a tandem architecture with two glycine binding domains adjacent to one another upstream of a single expression platform. The two ligand binding domains, which exhibit some distinct structural features, are referred to as Type I and Type II. The \textit{B. subtilis gcvT} leader sequence acts as a genetic “on” switch. When glycine is bound (as depicted), the terminator helix (potential pairing regions shown in red and blue) is destabilized resulting in transcription readthrough. Binding of glycine to the individual ligand binding domains is cooperative. Positioning of glycine within the ligand binding pockets has not yet been characterized by high resolution structure determination. Adapted from (Mandal et al., 2004).
1.2.5.2 L box riboswitch

The L box riboswitch is an RNA element that exhibits conserved structural motifs important for the recognition of the amino acid L-lysine. This riboswitch is located in genes involved in lysine biosynthesis and transport (Grundy et al., 2003; Sudarsan et al., 2003b), and modulates the levels of lysine available for translation and cell wall synthesis. It also regulates cellular pools of diaminopimelate (DAP), which is an intermediate in the lysine biosynthesis pathway and an important component of the cell wall in some bacterial species. The L box riboswitch has been identified in the genomes of many Firmicutes and Gammaproteobacteria, and in at least one species from the Order Thermotogales. Typically, L box elements encoded in the genomes of Gram-positive organisms utilize a mechanism of transcriptional control, while those encoded in the 5′UTRs of genes in Gram-negative organisms are predicted to use a mechanism of translational control.

The L box riboswitch is characterized by conserved secondary structure features and some primary sequence residues located in the helix junction region of the molecule. X-ray crystallography was used to develop a high resolution structural model of the L box riboswitch (Garst et al., 2008; Serganov et al., 2008). These data showed that lysine is encapsulated by the RNA and sits at the core of a 5-way junction, reminiscent of other riboswitches that also make use of a helical junction for ligand recognition. Lysine binding involves a potassium cation (K⁺) that makes a direct interaction with a lysine carboxyl oxygen and coordinates several indirect interactions (Serganov et al., 2008). The L box ligand binding domain binds lysine with an apparent $K_D$ of ~1 µM. However, the full-length leader RNA including the downstream regulatory element binds lysine less
efficiently with an apparent $K_D$ of $\sim 500$ µM (Sudarsan et al., 2003b). The lower binding affinity of the ligand binding domain alone is likely to be due to the loss of structural features required for formation of the unliganded conformation of the riboswitch RNA. Therefore, the $K_D$ of the full-length molecule is likely to be more physiologically relevant.

As with other riboswitches, the L box RNA is highly specific for its ligand and discriminates against closely related molecules including DAP and D-lysine (Grundy et al., 2003; Sudarsan et al., 2003b). Mutations at certain residues in the L box leader have been associated with resistance to the lysine analog aminoethycysteine (AEC), which inhibits growth of some organisms. However, the L box riboswitch binds AEC at least 10-fold less efficiently than lysine, suggesting that this analog does not act as a repressor of L box genes in vivo. Instead, the primary mode of action of AEC is probably through misincorporation of the analog by the lysyl-tRNA synthetase (LysRS) during translation (Ataide et al., 2007). Resistance in L box mutants is likely to be due to derepression of the downstream ORFs, leading to increased production of lysine, which is then able to outcompete AEC during tRNA aminoacylation.

1.2.6 glmS ribozyme

A ribozyme element that cleaves the mRNA element upstream of the glmS coding region has been identified in a number of Gram-positive bacteria (Winkler et al., 2004) (Fig. 1.5). The enzyme encoded by glmS, glutamine-fructose-6-phosphate amidotransferase, generates glucosamine-6-phosphate (GlcN6P), which is the molecular effector recognized by the glmS ribozyme. Investigations of the glmS leader sequence
from *B. subtilis* revealed that spontaneous site-specific RNA cleavage is accelerated 1000-fold in the presence of GlcN6P, but not in the presence of analogs. Cleavage of the *glmS* leader RNA results in decreased expression of the downstream coding sequence (Winkler et al., 2004). Mutation of conserved residues important for *glmS* ribozyme cleavage results in derepression of gene expression *in vivo*. It is postulated that the *glmS* ribozyme is prefolded prior to ligand binding and that GlcN6P is directly involved in the ribozyme cleavage event. This is in contrast to other known riboswitch classes in which ligand-dependent conformational changes are essential for riboswitch function. A modified ribozyme containing a 2′deoxy substitution at the nucleotide upstream of the cleavage site was used to investigate conformational changes in the RNA in response to GlcN6P (Hampel and Tinsley, 2006). This modification is sufficient to disrupt cleavage, while formation of the ribozyme structure is unperturbed. Data from these studies support the model that the *glmS* leader RNA does not undergo significant conformational changes in response to ligand binding.
Fig. 1.5 The GlcN6P responsive ribozyme in the \textit{glmS} gene in \textit{B. subtilis}. In the presence of GlcN6P, the leader RNA is site-specifically cleaved (arrow) and the resulting downstream fragment is targeted for degradation. The ligand binding pocket is situated directly in the center of the ribozyme active site, and residue G1, which is located just 3’ of the cleavage site, is required for ribozyme catalysis. The \textit{glmS} riboswitch is the only known riboswitch class that utilizes ribozyme-directed mRNA self-cleavage to regulate gene expression in response to ligand binding.

The \textit{glmS} cleavage reaction is predicted to proceed through a transesterification event in which the 2’OH of the nucleotide that lies immediately upstream of the cleavage site performs a nucleophilic attack on the adjacent phosphodiester (McCarthy et al., 2005). Several studies demonstrated that GlcN6P sits directly in the active site of the \textit{glmS} ribozyme in a solvent-accessible binding pocket, and acts as a coenzyme for enzyme activity in which the amine group plays a key role in proton transfer (Cochrane et al., 2009; Cochrane et al., 2007; Klein and Ferre-D'Amare, 2006; Klein et al., 2007b; McCarthy et al., 2005). Additionally, a conserved guanine residue has been implicated in ribozyme catalysis, and disruption of this nucleotide is sufficient to disrupt GlcN6P-dependent cleavage, even though ligand binding is unaffected (Klein et al., 2007a). Subsequent to cleavage by the \textit{glmS} ribozyme, the downstream mRNA is targeted for
degradation by ribonuclease J1 (RNase J1) (Collins et al., 2007). The \textit{glmS} element represents a unique riboswitch that uses a distinct mechanism of mRNA self-cleavage for modulating gene expression.

1.2.7 \textbf{Mg}^{2+} \textbf{sensing riboswitch}

The smallest known molecular effector that acts as a signal for riboswitch function is the divalent cation magnesium (Mg$^{2+}$). It is well-established that metal ions are essential for RNA folding (Draper, 2008). The rearrangement of RNA into a compact tertiary structure is complicated by the fact that the nucleic acid backbone exhibits a negative charge conferred by the repeating phosphate groups. This challenge is partially mitigated in the presence of positively charged ions such as Mg$^{2+}$ and K$^+$. Both of the known Mg$^{2+}$-responsive riboswitches exploit this property to adopt alternate conformations depending on the intracellular levels of magnesium.

The first Mg$^{2+}$-responsive riboswitch to be described is located in the leader region of the \textit{Salmonella enterica} \textit{mgtA} gene, which encodes a Mg$^{2+}$ transporter (Cromie et al., 2006). Although this RNA element does not appear to contain a typical intrinsic transcription terminator, it appears that regulation involves an attenuation mechanism in which high Mg$^{2+}$ levels result in decreased transcription of the \textit{mgtA} coding sequence. The second Mg$^{2+}$ riboswitch to be discovered (the M box) is located in ion transport genes in many Gram-positive bacteria, and exhibits no sequence similarity to the \textit{S. enterica} \textit{mgtA} riboswitch (Dann et al., 2007). Expression of M box-regulated genes is repressed via stabilization of an intrinsic terminator helix when Mg$^{2+}$ levels are sufficiently high. The M box element in the \textit{mgtE} gene in \textit{B. subtilis} was shown to bind
Mg\(^{2+}\) with an apparent $K_D$ of ~2.6 mM. Determination of the crystal structure of the M box element in the *B. subtilis* Mg\(^{2+}\) transport gene *ykoK* revealed that the binding pocket lies at the convergence of three parallel helices, and that six Mg\(^{2+}\) ions are important for mediating long-range interactions.

The specificity of each of the Mg\(^{2+}\) riboswitches has been demonstrated *in vivo* using a variety of divalent cations (Cromie et al., 2006; Dann et al., 2007). Addition of alternate metal ions, at levels that are sufficient for conferring Mg\(^{2+}\)-dependent repression, did not result in decreased expression of reporter gene constructs. While some structural rearrangement of the M box riboswitch has been observed in response to high concentrations (5 mM) of calcium (Ca\(^{2+}\)) and manganese (Mn\(^{2+}\)) *in vitro*, it is not likely that these cations are freely available at levels sufficient to confer regulatory control *in vivo*. Therefore, the M box riboswitch appears to be specific and responds only to the magnesium pools inside the cell. While these are the only two riboswitch classes that have been shown to respond directly to divalent cations, other elements that may act as metalloregulatory RNAs have been identified in genes involved in metal ion homeostasis (Dann et al., 2007).

### 1.2.8 FMN riboswitch

A conserved motif referred to as the RFN element has been identified in genes involved in riboflavin biosynthesis in some bacteria (Mironov et al., 2002; Winkler et al., 2002b). Riboflavin (vitamin B\(_2\)) is an important cellular cofactor, and the metabolically active form (flavin mononucleotide, FMN) has been shown to interact directly with the RFN element to confer regulatory control of the biosynthetic genes. Determination of the
crystal structure of the RFN riboswitch revealed the overall architecture of the ligand binding domain, which consists of a six-helix junction enclosing the FMN molecule (Serganov et al., 2009). The RFN element binds FMN with an apparent $K_D$ of $\sim 0.04 \mu M$, and analogs with altered functional groups are recognized with lower affinity. For example, the naturally occurring precursor riboflavin, which lacks the phosphate group, binds the RFN element with 1,000-fold less affinity (apparent $K_D \sim 40 \mu M$). The crystal structure revealed that the phosphate oxygens of FMN form hydrogen bond interactions with the Watson-Crick faces of conserved guanine residues in the RNA ligand binding pocket. These interactions require coordination of a hydrated Mg$^{2+}$ ion, reminiscent of other riboswitch systems that also preferentially bind phosphate-containing ligands (e.g., glmS and THI box riboswitches). Another naturally occurring analog, flavin adenine dinucleotide (FAD), consists of FMN covalently attached to an adenine via its phosphate moiety. This molecule has been shown to promote transcription termination of the RFN element in vitro, although a 17-fold higher concentration was required to elicit a response similar to that observed with FMN (Mironov et al., 2002). These results suggest that FMN is likely to act as the primary effector molecule in vivo. Two recent studies have suggested that the naturally occurring riboflavin analog roseoflavin, produced by Streptomyces davawensis, may target the FMN riboswitch in B. subtilis (Lee et al., 2009; Ott et al., 2009). These results help to explain the mild antimicrobial activity of this molecule against several bacterial species, and provide an example of how riboswitches might serve as effective targets for drug development.

An investigation of the RFN element residing upstream of the ribD coding region in B. subtilis revealed that this riboswitch element uses a mechanism of kinetic
control to regulate gene expression (Wickiser et al., 2005). In this system, the efficiency of gene expression is not determined by the equilibrium dissociation constant ($K_D$), since during the course of transcription, the RNA-ligand complex does not have sufficient time to reach thermodynamic equilibrium before the RNA polymerase (RNAP) has to commit to a regulatory decision (i.e., readthrough or terminate). These results help to explain why more ligand (~30 to 100-fold) is required to reach half-maximal transcription termination in an in vitro assay than to reach half-maximal binding at equilibrium. This riboswitch relies instead on the relative rates of RNA polymerase transcription and association of the FMN effector molecule with the leader RNA to determine whether the downstream coding sequence is expressed. The RNA element used in these studies was shown to contain pause sites upstream of the regulatory domain that are postulated to stall transcription to allow more time for FMN association. Results from these studies also demonstrated that the FMN-RNA complex has a half-life ($t_{1/2}$; time required for half of the complex to dissociate) in excess of 10 min. This is predicted to be longer than the typical half-life of many mRNAs inside the cell (Selinger et al., 2003), and also longer than the time required for RNAP to transcribe through the terminator helix. This suggests that regulation by this riboswitch is irreversible. The mechanism of kinetic control described here may represent a common mechanism used by riboswitch classes that regulate at the level of premature transcription termination, in which ligand binding and mRNA synthesis occur simultaneously. Not all FMN riboswitches use a mechanism of transcriptional control. For instance, the RFN element in the 5’ UTR of the ypaA gene in B. subtilis is predicted to regulate at the level of translation initiation.
1.2.9 B\(_{12}\) riboswitch

Another enzyme cofactor that acts as a molecular effector for riboswitch activity in some microorganisms is 5’deoxy-5’-adenosylcobalamin (AdoCbl), or vitamin B\(_{12}\). The first B\(_{12}\) riboswitch to be characterized is located in the 5’UTR of the \(btuB\) gene in \(E.\ coli\) and \(Salmonella\ typhimurium\), which encodes a cobalamin transport protein (Nahvi et al., 2002; Nou and Kadner, 2000). This riboswitch element specifically binds AdoCbl with an apparent dissociation constant of \(~300\ \text{nM}\), and undergoes a structural rearrangement in response to ligand that results in sequestration of the RBS. AdoCbl represents one of the largest metabolites that acts as an effector for riboswitch function, and includes a variety of functional groups that may serve as hydrogen bond donor/acceptors. While no high resolution structure is currently available for the AdoCbl binding domain, experiments with analog compounds have identified ligand moieties that are important for recognition (Nahvi et al., 2002). For instance, modification of the adenine ring of AdoCbl disrupts ligand binding, while modification of the 2’hydroxyl on the adjoining ribose does not. The B\(_{12}\) riboswitch has been identified in the genomes of many Gram-positive and Gram-negative organisms including some human pathogens (Nahvi et al., 2004; Warner et al., 2007), and has in some instances been shown to utilize a regulatory mechanism of premature transcription termination.

The B\(_{12}\) riboswitch has been observed in the context of two different types of tandem arrangements (Sudarsan et al., 2006), in which two riboswitch elements are situated adjacent to one another upstream of a single ORF. The simpler of the two arrangements involves two identical B\(_{12}\) responsive riboswitches that both contain an intrinsic transcription terminator motif, and that act independently to regulate expression
of the DSY1435 mRNA in *Desulfobacterium hafniense*. Two independent riboswitches acting in tandem are postulated to allow for greater responsiveness to subtle changes in ligand levels. However, this type of arrangement contrasts with the tandem architecture observed with the glycine riboswitch (described in section 1.2.5.), in which ligand binding is cooperative, and the presence of glycine at one RNA element increases the affinity for glycine at the other element (see section 1.2.5.1; Mandal et al., 2004).

The second type of tandem arrangement, which has been identified in the 5′ UTR of the *metE* gene in *Bacillus clausii*, is comprised of an S-adenosylmethionine (SAM)-responsive S box riboswitch followed by a B12 riboswitch (Fig. 1.6). MetE is one of two isoenzymes responsible for the formation of methionine from homocysteine. The alternate enzyme, MetH, requires AdoCbl for the synthesis of methionine while MetE does not require AdoCbl. Both of the genes encoding these enzymes are regulated by SAM binding S box riboswitches, but only *metE* is also regulated by a B12-responsive element. Therefore, when AdoCbl levels are high the *metE* gene is turned off allowing for the preferential use of the more efficient protein MetH for methionine synthesis. Alternately, when SAM levels are high, both genes are downregulated via the S box mechanism. The tandem riboswitches that reside in the *metE* 5′UTR both contain intact terminator helices, and have been shown to operate independently (Sudarsan et al., 2006). The presence of either molecular effector (SAM or AdoCbl) is sufficient to confer repression, and the binding affinity for one ligand is unaffected by the presence of the other.
Fig. 1.6. Tandem S box and B12 riboswitches in the 5’UTR of the *metE* gene in *Bacillus clausii*. The leader sequence of the *metE* gene contains an S box riboswitch immediately upstream of a B12 riboswitch. Both elements include a separate ligand binding domain and expression platform such that each riboswitch operates independently. Binding of either molecular signal, SAM (*) or AdoCbl (*), is sufficient to promote formation of a terminator helix (T) and repress expression of the downstream coding sequence. Localization of AdoCbl in the B12 riboswitch ligand binding pocket has not yet been characterized. Adapted from (Sudarsan et al., 2006).

1.2.10 THI box riboswitch

One of the earliest discovered riboswitches is the THI box riboswitch (Miranda-Rios, 2007). This RNA element binds the enzymatic cofactor thiamin pyrophosphate (TPP), which is the biologically active form of thiamin (vitamin B1). The THI box riboswitch is the most widespread of all the known riboswitch classes, and is the only riboswitch that has been observed in all three domains of life (Sudarsan et al., 2003a). Similar to many other riboswitch classes, the THI box element is modular in architecture, and although the ligand binding domain exhibits highly conserved structural motifs, the regulatory domain displays a high level of structural and mechanistic diversity in
different organisms (Bocobza and Aharoni, 2008). Additionally, this riboswitch has been observed in some organisms in a tandem architecture in which two THI box elements, complete with ligand binding domains and regulatory output domains, operate independently to confer enhanced responsiveness to changing TPP levels (Sudarsan et al., 2006; Welz and Breaker, 2007).

In bacteria, the THI box RNA is found in the 5’UTR of genes involved in the biosynthesis and transport of thiamin, and regulates expression of downstream coding sequences through either premature transcription termination or translation inhibition upon ligand binding (Miranda-Rios et al., 2001; Winkler et al., 2002a). In some plants, the THI box consensus sequence has been identified in the 3’UTR of the thiamin biosynthetic gene thiC where it regulates mRNA processing and stability (Bocobza et al., 2007; Wachter et al., 2007). TPP binding at a THI box element located in an intron downstream of the thiC gene in Arabidopsis thaliana results in increased splicing, and production of a transcript with decreased stability relative to the unprocessed transcript. TPP-responsive RNAs that have been identified in the genomes of some filamentous fungi and green algal species also use a regulatory mechanism that involves splicing. However, in these latter organisms the THI box consensus sequence is typically located in the 5’ region of the mRNA. TPP binding to the THI box riboswitch located in the NMT1 gene in the filamentous fungi Neurospora crassa leads to increased production of an alternatively spliced product that contains upstream ORFs (uORFs) that compete for translation initiation and thereby repress expression of the main ORF (Cheah et al., 2007). In the green alga Chlamydomonas reinhardtii, THI box elements have also been shown to affect splicing of introns containing competing uORFs (e.g., the THI4 gene), or
in some instances introns containing in-frame stop codons (e.g., the THIC gene) that affect the efficiency of gene expression (Croft et al., 2007). In each of the eukaryotic THI box mechanisms described, TPP binding is thought to increase the accessibility of alternative splice sites through allosteric modulation of the RNA structure. The remarkable versatility of regulatory strategies employed by the THI box riboswitch in various contexts demonstrates how RNA can adapt alternate strategies for altering gene expression.

Recently, crystal structures of the *A. thaliana* thiC and *E. coli* thiM riboswitches in complex with ligand were reported, providing crucial insight into ligand-RNA interactions and tertiary folding (Edwards and Ferre-D'Amare, 2006; Serganov et al., 2006; Thore et al., 2008; Thore et al., 2006). Despite the different regulatory mechanisms utilized by these two riboswitches, the structures of the ligand binding domains are virtually identical. The THI box riboswitch forms an overall structure that resembles a tuning fork comprised of two parallel helical domains (P2/P3 and P4/P5) adjoined to a third helix (P1) via a 3-way junction. The two parallel helices of the RNA also interact with each other when ligand is bound through contacts between the terminal loop of P5 and nucleotides in the P3 helix domain. TPP is positioned between the two parallel helices in an extended conformation, and makes key contacts via its pyrimidine ring (with the P2/P3 helix) and pyrophosphate group (with the P4/P5 helix). The requirement for the ligand to assume an extended conformation explains the decreased binding affinity of the RNA for the analogs thiamin monophosphate (TMP) and thiamin, which exhibit decreased molecular width relative to TPP. The binding affinities for each of these ligands is approximately 0.1 µM for TPP, 100 µM for TMP, and 600 µM for thiamin.
(Winkler et al., 2002a), although reported values are somewhat variable depending on the methodology and the length of THI box sequence used. The crystal structures also reveal that the central thiazole ring of TPP is not recognized, which provides a rationale for the binding activity of the antibiotic pyrithiamine pyrophosphate (PTPP), which possesses a central pyridine ring in place of the thiazole ring (Thore et al., 2008). Another important feature of the THI box riboswitch is the requirement for water-coordinated divalent cations that help to counter the negative repulsion forces between the pyrophosphate moiety of TPP and the RNA backbone. The use of divalent cations in this manner, which also has been observed in other riboswitch systems, expands the repertoire of ligand functional groups with which RNAs can interact.

The *E. coli thiM* riboswitch controls gene expression at the level of translation initiation (Fig. 1.7). When TPP levels are low, the AASD sequence pairs with the downstream ASD sequence. In this conformation, the SD sequence, which is further downstream, is accessible for interaction with the 30S ribosomal subunit and translation initiation. Binding of TPP stabilizes the helix at the base of the ligand binding pocket (helix P1) so that the AASD sequence is no longer available for pairing with the ASD sequence. Under these conditions, the RNA favors a conformation in which the ASD sequence pairs with the SD sequence, thereby preventing binding of 30S ribosomal subunits to the TIR. This model is supported by phylogenetic and structural probing data (Rentmeister et al., 2007; Winkler et al., 2002a). Chapter 2 will describe work with the *E. coli thiM* leader RNA to investigate the effects of TPP on SD sequence accessibility and 30S ribosomal subunit binding.
Fig. 1.7. THI box riboswitch in the 5′UTR of the thiM gene in E. coli. The E. coli THI box element regulates at the level of translation initiation. In the absence of TPP (left), the RNA favors a conformation in which the SD sequence (green) is accessible for binding to the translation initiation complex. In this conformation, the ASD sequence (red) is sequestered by an upstream sequence called the AASD sequence (turquoise). When TPP levels are high (*, right), the RNA favors a conformation in which the AASD sequence is sequestered at the base of the ligand binding domain, thereby freeing up the ASD sequence for interaction with the SD sequence. In this conformational state, interaction with the 30S ribosomal subunit is prevented and translation initiation is inhibited. The AUG start codon is indicated in yellow.

1.2.11 SAM riboswitches

There are three known riboswitch classes (S box/SAM-I, SAM-II, and S_{MK} box) that bind to the cellular cofactor S-adenosylmethionine (SAM; McDaniel et al., 2003; Corbino et al., 2005; Fuchs et al., 2006; Wang and Breaker, 2008). Each of these RNAs employs a distinct architecture for recognition of the same ligand, and discriminates against the naturally occurring analog S-adenosylhomocysteine (SAH), which differs from SAM by a single methyl group and a positive charge on the sulfur. A fourth riboswitch class has been described that specifically recognizes SAH, and discriminates
instead against SAM. The SAM and SAH binding riboswitches are typically associated
with genes involved in methionine and cysteine metabolism, and although a single
riboswitch class can be involved in the regulation of multiple transcriptional units in a
single organism (e.g., the S box riboswitch), different classes are not typically found in
the same species. The structural diversity among these riboswitch classes suggests that
each of these RNA elements originated from an independent evolutionary pathway, and
demonstrates the versatility of RNA as a molecular sensor. The following sections will
describe what is currently known about each of these riboswitch classes.

1.2.11.1 S box riboswitch

The S box riboswitch was first identified in the leader sequences of 11 different
transcriptional units in *B. subtilis*, and in the genomes of several other Gram-positive
bacteria (Grundy and Henkin, 1998). It was observed that expression of the S box
regulated gene *yitJ* is repressed during growth in the presence of methionine, and that
regulation is dependent on an intrinsic transcription terminator located in the leader
sequence. Secondary structural features located upstream of the terminator helix were
predicted to constitute the binding site for a regulatory factor. The S box RNA was shown
to interact directly with SAM, and to undergo a conformational change in response to
SAM binding that enhances transcription termination *in vitro* in the absence of protein
factors or translating ribosomes (McDaniel et al., 2003). These results support the model
that in the absence of SAM the S box element favors a conformation in which an
antiterminator element is formed and gene expression is turned on (Fig. 1.8). When SAM
levels are high, the RNA undergoes a structural rearrangement that results in formation of
anti-antiterminator element, which sequesters a portion of the antiterminator and frees up
nucleotides required for formation of the terminator helix. This model is supported by results obtained from mutational analyses and structural probing experiments (Epshtein et al., 2003; McDaniel et al., 2003; Winkler et al., 2003).

**Fig. 1.8. S box mechanism of transcriptional control.** The predicted secondary structure depicted here is based on the *yitJ* leader sequence from *B. subtilis*. In the absence of SAM (left), a stable antiterminator (AT) element is formed and transcription proceeds into the downstream coding region. In the presence of SAM (*, right), residues required for formation of the antiterminator element are sequestered in an alternate helix (P1/AAT), allowing formation of the terminator helix (T). Tertiary interactions between the loop of P2 and the P3/P4 junction in the ligand-bound conformation are depicted by a dashed line. Adapted from (Tomsic et al., 2008).

The S box riboswitch is widespread relative to other SAM binding riboswitches, and employs several different regulatory mechanisms (Rodionov et al., 2004). It is most common in Firmicutes where it regulates at the level of transcription termination, but it has also been identified in some Gram-negative organisms and Actinomycetes where it is predicted to regulate at the level of translation initiation. In the case of *Clostridium*
acetobutylicum, an S box riboswitch is predicted to regulate expression of an antisense RNA with complementarity to an mRNA that encodes proteins involved in the conversion of methionine to cysteine. In this system, when SAM levels are low the full-length antisense RNA is transcribed, and expression of the methionine recycling genes is repressed allowing for accumulation of methionine. Finally, as noted above (see section 1.2.9), an S box element has been observed in tandem with a vitamin B\textsubscript{12}-responsive riboswitch in the 5’UTR of the metE gene in B. clausii. In this context, both riboswitches operate independently and either regulatory signal (SAM or AdoCbl) is sufficient to repress expression of the downstream coding sequences (Sudarsan et al., 2006).

A secondary structure model was formulated based on the phylogenetic analysis of S box sequences from B. subtilis and other organisms (Grundy and Henkin, 1998). According to this model, the S box element consists of four helical regions (P1 to P4) that converge in a 4-way junction to form the SAM binding pocket. In the ligand-bound conformation, helix P1 is stabilized and a portion of the sequence that forms the antiterminator element is sequestered in an anti-antiterminator element, allowing formation of the downstream terminator helix. In the absence of SAM, P1 is destabilized, and residues on the 3’ side are involved in the formation of the antiterminator element, which disrupts formation of the mutually exclusive terminator structure. Tertiary interactions between the loop of P2 and the junction of P3 and P4 were predicted based on phylogenetic analysis (Grundy and Henkin, 1998) and confirmed by mutational analysis and structural probing experiments (McDaniel et al., 2005). A kink-turn motif in helix P2 was also predicted to be important for facilitating tertiary interactions in the S box structure (Winkler et al., 2001).
These predictions were validated by determination of the crystal structure of an S box riboswitch from *Thermoanaerobacter tengcongensis* in complex with SAM (Montange and Batey, 2006). The crystal structure revealed that the riboswitch assumes a global fold comprised of two sets of coaxial stacked helices (P1/P4 and P2/P3) with SAM positioned between helices P1 and P3. The SAM molecule is enveloped on virtually all sides and assumes a compact conformation in which the methionine moiety stacks on top of the adenine. The methyl group of SAM is not directly recognized and some substitutions at this moiety are tolerated for ligand recognition. The positive charge on the sulfur group of SAM makes favorable electrostatic interactions with the carbonyl oxygens of two conserved uridine residues in the RNA. This latter feature explains the preferential binding of SAM over SAH (by at least ~100-fold), and is observed in other classes of SAM binding riboswitches.

It was recently demonstrated that the S box elements in *B. subtilis* exhibit differential sensitivity to SAM in terms of ligand binding affinity and regulatory response (Tomsic et al., 2008). This suggests that riboswitch function can be fine-tuned through minor differences in RNA structure to modulate expression of individual genes. These studies also revealed that the S box regulatory elements exhibit a SAM-S box complex half-life that is much longer than the time required for transcription to reach the terminator helix. This indicates that the S box in *B. subtilis* is likely to be an irreversible riboswitch, and that once SAM binding occurs, the messenger RNA is committed to premature transcription termination. As with other riboswitches that are regulated by transcription attenuation (e.g., the FMN riboswitch), the kinetics of mRNA transcription and ligand association may be crucial for determining the regulatory outcome.
1.2.11.2 SAM-II riboswitch

The second SAM-responsive riboswitch to be discovered was the SAM-II element that is found primarily in Proteobacteria (Corbino et al., 2005). The initial characterization of this regulatory RNA was done on the Agrobacterium tumefaciens metA leader RNA, which was shown to bind SAM with an apparent $K_D$ of ~1 µM. The SAM-II riboswitch is much smaller in size than the S box riboswitch, and it assumes a very different structural arrangement for formation of the SAM binding pocket. The crystal structure of a SAM-II element obtained from an environmental sample revealed that this RNA forms a ligand-dependent H-type pseudoknot structure that results in occlusion of the SD sequence (Gilbert et al., 2008). In contrast to the S box system, the SAM molecule in complex with the SAM-II riboswitch assumes an extended conformation. However, similar to the S box system, the positive charge on the sulfur group of SAM is essential for maintaining electrostatic interactions with conserved uridines in the RNA. All of the SAM functional groups are recognized by the SAM-II mRNA, and substitutions at the methyl position are not tolerated as they are in the S box system (Corbino et al., 2005; Gilbert et al., 2008). Unlike most other classes of riboswitches, the regulatory domain of the SAM-II riboswitch is predicted to be intrinsic to the ligand binding domain, although the regulatory mechanism has not been demonstrated. This type of simple architecture has been observed in only one other riboswitch class, the S_{MK} box.

1.2.11.3 S_{MK} box riboswitch

An investigation of the genomes of Gram-positive lactic acid bacteria revealed that the metK gene (which encodes S-adenosylmethionine synthetase, the enzyme
responsible for the synthesis of SAM from methionine and ATP) is not regulated by an S box (SAM-I) or SAM-II riboswitch. Instead, a new sequence motif was discovered that constitutes the third class of SAM binding riboswitch, and is referred to as the $S_{MK}$ box or SAM-III riboswitch (Fuchs et al., 2006). This riboswitch has been found only in $metK$ genes in most members of the Lactobacillales. All $S_{MK}$ box elements that have been identified are predicted to regulate at the level of translation initiation through partial sequestration of the SD sequence by an upstream ASD sequence (Fig. 1.9). In the absence of SAM, the ASD-SD interaction is predicted to be disrupted, thereby freeing up the SD sequence for interaction with the 30S ribosomal subunit. Fusions of the $Enterococcus faecalis$ $metK$ leader sequence to a $lacZ$ reporter demonstrated that the $S_{MK}$ box riboswitch regulates at the level of translation, and the structural model was further supported by data obtained using structural probing and mutational analysis (Fuchs et al., 2006). Also, it has been directly demonstrated that addition of SAM to $S_{MK}$ box RNA leads to decreased interaction with 30S ribosomal subunits in vitro (Fuchs et al., 2007), providing strong evidence in support of the model.
Fig. 1.9. Model of the *E. faecalis* $S_{MK}$ box mechanism. Based on phylogenetic analysis, it was predicted that the $S_{MK}$ box riboswitch regulates at the level of translation initiation in response to SAM ($^*$). In the absence of SAM (left), the riboswitch favors a conformation in which the SD sequence (green) is accessible. SAM binding (right) results in a structural rearrangement that leads to occlusion of the SD sequence by pairing with an upstream ASD sequence (red). The AUG start codon is indicated in yellow. Modified from (Fuchs et al., 2006).

The $S_{MK}$ box motif is characterized by highly conserved residues within the core of the RNA that have been shown to be important for SAM binding. It was predicted that many of these residues are involved in maintaining the overall structure of the ligand binding pocket or in direct interactions with SAM. Based on phylogenetic analysis, a secondary structure model was developed in which the RNA is predicted to assume a conformation involving the formation of several helical regions (Fuchs et al., 2006). Sequence comparisons revealed two hypervariable regions in which the number of intervening nucleotides and the predicted primary and secondary structures differ
dramatically between species. This feature of the $S_{\text{MK}}$ box RNA can make identification of this element problematic since the distances between conserved structural features are so variable. Indeed, the $S_{\text{MK}}$ sequences identified to date range in size from ~80 nt (as in *Enterococcus faecium*) to over 380 nt (as in *Streptococcus pyogenes*). In addition to the SAM-II riboswitch, the $S_{\text{MK}}$ box is the only riboswitch characterized to date that is not modular in architecture, i.e., the regulatory domain is intrinsic to the ligand binding domain. Chapters 3 and 4 of this document will outline structural studies that investigate ligand binding to the $S_{\text{MK}}$ box riboswitch, and Chapter 5 will outline fluorescence studies that investigate the affinity, specificity, and kinetics of ligand binding.

### 1.2.11.4 SAH riboswitch

A riboswitch that specifically recognizes SAH has recently been characterized (Wang et al., 2008). This riboswitch element is located in the 5’UTR of genes involved in the conversion of SAH to methionine in many proteobacteria and in some actinobacteria. The gene products regulated by the SAH riboswitch are important for maintaining sufficient intracellular levels of SAM, and also for recycling SAH, which can become toxic at high levels. Therefore, the SAH riboswitch is predicted to utilize a regulatory mechanism in which gene expression is induced in response to ligand binding, using either a mechanism of transcriptional or translational control.

Based on the secondary structure model, the SAH riboswitch is predicted to form several helical domains (P1, P2, and P4), with a pseudoknot element that is predicted to be stabilized upon ligand binding (Wang et al., 2008). This riboswitch also exhibits an additional helical domain (P3) that is present only in some species. The 68-nt SAH riboswitch in the leader sequence of the *metH* gene in *Dechloromonas aromatica*
binds SAH with an apparent $K_D$ of ~20 nM, and binds SAM with 1,000-fold lower affinity. Experiments using SAH analogs demonstrated that most of the SAH functional groups are critical for ligand recognition, although alterations at the methionine tail are somewhat tolerated. For example, the analog S-adenosylcysteine (SAC), which contains a shorter amino acid side chain, resulted in only a ~10-fold reduction in binding affinity. This is in contrast to the SAM binding riboswitches, which are more sensitive to changes at this functional group. Determination of a high resolution structure of the SAH riboswitch will provide insight into how this RNA specifically recognizes its ligand, and will allow for direct comparison to the three known classes of SAM binding riboswitches.

1.3 Project Goals

My early research focused on investigation of the mechanism of regulation employed by the THI box riboswitch located in the 5’UTR of the $thiM$ gene in $E. coli$ (see Chapter 2). These studies were carried out in collaboration with the laboratory of Dr. Juan Miranda-Ríos at the Universidad Nacional Autónoma de México. While previous studies have shown that the riboswitch undergoes a conformational change in response to TPP (Mironov et al., 2002; Winkler et al., 2002a), there has been no direct demonstration of the effect of ligand binding on accessibility of the SD sequence. We demonstrated that the SD sequence is in fact occluded in the presence of TPP, and that 30S ribosomal subunit binding is subsequently inhibited (Ontiveros-Palacios et al., 2008). These data provided direct evidence for the translational control mechanism. We also tested a variety of constructs containing mutations in the ligand binding domain, to investigate the effect of those mutations on the regulatory output domain located downstream. These
experiments identified mutations that result in a loss of TPP binding, and subsequent loss of repression. In contrast, a second group of mutant constructs were identified that result in constitutive repression even in the absence of ligand. These data provided insight into the mechanism of TPP-dependent regulation and the interplay between the ligand binding domain and the regulatory output domain in a modular riboswitch system.

The \( \text{SMK} \) box riboswitch regulates gene expression at the level of translation through the SAM-dependent sequestration of the SD sequence. Previous investigations supported this mechanism (Fuchs et al., 2006, 2007). However, limited information was available regarding the overall structure of the \( \text{SMK} \) box riboswitch and the mechanism by which the RNA recognizes SAM. A primary focus of my research has been to investigate the RNA structural requirements for SAM binding. To this end, an extensive mutational analysis has been performed (Lu et al., 2008, A.M. Smith and R.T. Fuchs, unpublished) to examine the importance of conserved residues using an \textit{in vitro} SAM binding assay (see Chapter 3). We demonstrated that many of the conserved residues located in the core of the \( \text{SMK} \) box RNA are essential for SAM binding. In order to determine the precise roles of individual nucleotides, we used a minimal \( \text{SMK} \) box construct (truncated to facilitate crystallization) to determine the 2.2-Å resolution crystal structure of the RNA-SAM complex, in collaboration with the laboratory of Dr. Ailong Ke at Cornell University (Lu et al., 2008). These experiments provided detailed information about key nucleotides involved in direct contacts with SAM as well as structural features important for stabilizing the ligand binding pocket and discriminating against closely related compounds such as SAH. The structural data provided insight into the importance of the
SD sequence for ligand binding, and allowed a direct comparison to the S box and SAM-II riboswitches.

Results from the mutational analysis and x-ray crystallography experiments allowed us to garner information regarding structural features of the SAM-bound SMK box RNA. However, we were also interested in characterizing the alternate conformation adopted by the RNA in the absence of SAM. To address this aim, we used Nuclear Magnetic Resonance (NMR) spectroscopy to investigate the secondary structure of the SMK box RNA (in collaboration with the laboratory of Dr. Mark Foster, The Ohio State University) in both conformational states (see Chapter 4). These experiments validated the secondary structure in solution, both in the presence and absence of SAM, and provided insight into the spatial proximity of individual nucleotides in the three-dimensional structure. In addition, these studies demonstrated that nucleotides immediately upstream of the ASD sequence are crucial for stabilization of the unliganded SMK box conformation, in which the SD sequence is available for translation initiation.

Another focus of my research has been the development of a fluorescence assay for investigating ligand binding to the SMK box riboswitch (see Chapter 5). Previously, it was demonstrated that the SMK box riboswitch specifically binds SAM in a competition binding experiment using SAH as a competitor (Fuchs et al., 2006). We used a fluorescence based assay to test binding of the \textit{E. faecalis} met\textit{K} leader RNA to various SAM-related compounds to further characterize the specificity of ligand binding (A.M. Smith, unpublished). These experiments provided insight into the functional groups of SAM that are important for ligand recognition, and into the ability of the SMK box riboswitch to discriminate against SAM analogs. We also used the fluorescence assay to
determine the rates of association and dissociation of the RNA-SAM complex. Our results established that the half-life of the S\textsubscript{MK} box RNA inside the cell is much longer than the half-life of the complex \textit{in vitro} (R.T. Fuchs and A.M. Smith, unpublished). These results indicate that SAM-dependent repression of the S\textsubscript{MK} box riboswitch may be reversible, and that a single RNA may undergo several regulatory decisions during the course of its lifetime. This is the first demonstration of the potential reversibility of a riboswitch RNA.
CHAPTER 2

MECHANISTIC STUDIES OF THE THI BOX RIBOSWITCH

2.1 Introduction

The THI box riboswitch is the most widespread riboswitch discovered to date (Miranda-Rios, 2007). It is the only riboswitch that has been identified in the genomes of representative organisms from all three domains of life. In addition to its prevalence, it also displays the most diversity in terms of the types of mechanisms it uses for conferring regulatory control. As a riboswitch system that assumes a modular architecture, the THI box element exhibits conserved structural features that comprise the ligand binding domain, and couples these motifs with a wide range of regulatory output domains that vary in both structure and function. In bacteria, the THI box riboswitch commonly utilizes a mechanism of regulatory control that involves inhibition of translation initiation or premature transcription termination, similar to many other characterized riboswitch classes. However, in eukaryotes the THI box riboswitch has been shown to utilize a mechanism of regulatory control involving alternative splicing in response to ligand binding that affects either mRNA stability or translation efficiency (see section 1.2.10). Although unprecedented in other known classes of riboswitches, the mechanistic diversity of the THI box system underscores the potential of RNA as a versatile
regulatory molecule, and raises the possibility for other alternative mechanisms across a wide range of species.

The THI box riboswitch binds the enzymatic cofactor thiamin pyrophosphate (TPP). TPP is the biologically active form of thiamin (vitamin B₁), and is comprised of a pyrimidine group, a central thiazole ring, and a pyrophosphate group (Fig. 2.1). The THI box RNA usually is located in genes involved in thiamin biosynthesis and transport (Miranda-Rios et al., 2001; Sudarsan et al., 2003a). One THI box element that has been the focus of several biochemical and structural studies is the leader sequence that is located in the 5’UTR of the \textit{thiM} gene in \textit{E. coli}. This riboswitch is predicted to regulate at the level of translation initiation (Fig. 2.1). In the TPP-bound form, the RNA folds into a structure in which the SD sequence is sequestered through base pairing with an ASD sequence and cannot be accessed by the ribosome. When TPP levels are sufficiently low, the ASD sequence forms an alternate helical structure with an upstream sequence, leaving the SD region accessible for binding to the translation initiation complex. This model is supported by data obtained using enzymatic and chemical probing (Rentmeister et al., 2007).
Fig. 2.1. Chemical structure of TPP and secondary structure of the THI box element in the 5’UTR of the thiM gene in E. coli. A) The chemical structure of thiamin pyrophosphate is depicted with the pyrimidine ring on the left, the thiazole group in the middle, and the pyrophosphate group on the right. B) The THI box riboswitch that is located in the 5’UTR of the thiM gene in E. coli is predicted to regulate at the level of translation initiation. In the absence of TPP (left), the RNA assumes a conformation in which the SD sequence (green) is accessible. In this conformation, the ASD sequence (red) is sequestered by the AASD sequence (turquoise) upstream. When TPP binds (right), the RNA undergoes a structural rearrangement in which the AASD sequence is sequestered in a helix at the base of the ligand binding domain. In this conformation the ASD sequence is free to base pair with the SD sequence and expression of the downstream coding sequence is repressed. The thiM AUG start codon is indicated in yellow.
High resolution crystal structures of THI box ligand binding domains from *A. thaliana* and *E. coli*, in complex with TPP, have recently been published (Edwards and Ferre-D'Amare, 2006; Serganov et al., 2006; Thore et al., 2006). These structural data revealed a high level of similarity between the two RNAs, providing strong evidence for evolutionary conservation of the riboswitch across diverse organisms. In both structures, the THI box ligand binding domain is comprised of five helical regions (P1-P5) that assume an overall tertiary structure that resembles a tuning fork (Fig. 2.2A and B). Two parallel helices, consisting of P2 coaxially stacked with P3, and P4 stacked with P5, are connected to helix P1 via a centrally located three-way junction. TPP binds between the two coaxially stacked helices with its pyrimidine moiety contacting a bulge in the P2/P3 region, and its pyrophosphate group interacting with the P4/P5 helix. Ligand interaction requires the use of divalent cations to counter repulsive forces between the negatively charged pyrophosphate group of TPP and the phosphate backbone of the riboswitch RNA (Edwards and Ferre-D'Amare, 2006). Although these studies provide crucial insight into the structural features that make up the ligand binding domain of the THI box riboswitch, they fail to provide a comprehensive picture of the complete system that includes the expression platform downstream.

A previous study using chemical and enzymatic probing revealed that the pyrimidine binding helical domain (P2/P3) of the THI box riboswitch is preformed in the absence of TPP, while the pyrophosphate binding moiety is not (Rentmeister et al., 2007). Addition of TPP results in organization of the pyrophosphate binding domain, and the concomitant formation of helix P1 at the base of the ligand binding pocket, which sequesters the AASD sequence (Fig. 2.1). This event renders the ASD sequence available.
for interaction with the SD sequence and gene expression is repressed. While these data provide insight into the interplay between the THI box ligand binding domain and expression platform, no direct experimental validation has been presented to show that 30S ribosomal subunit binding is diminished in response to TPP.

![Figure 2.2](image)

**Fig. 2.2. Structure of the *E. coli* THI box ligand binding domain in complex with TPP.** **A**) Secondary structure of the ligand binding domain in the presence of TPP (*). Residues that are conserved in >97% of known THI box sequences are shown in gray boxes (Abreu-Goedger et al., 2004; Ontiveros-Palacios et al., 2008). The AASD sequence, which is sequestered in the P1 helix upon binding of TPP, is indicated in turquoise. **B**) Crystal structure of the THI box riboswitch in complex with ligand. The overall structure of the THI box ligand binding domain resembles a tuning fork. TPP is shown encapsulated in the core of the RNA with magnesium ions represented by green spheres. The AASD sequence is indicated in turquoise as in A). Image was generated from coordinates deposited in the RCSB Protein Data Bank under accession number 2HOJ.
In the current study, an in vivo expression analysis using translational fusions confirmed that mutation of residues predicted to be involved in ligand binding resulted in a loss of TPP-dependent repression (work done by the laboratory of Dr. Juan Miranda-Rios at the Universidad Nacional Autónoma de México). Most of the mutants tested resulted in constitutive expression regardless of the presence of thiamin. However, a few of the mutants exhibited repression even during growth in the absence of thiamin, indicating that these constructs favor a conformation in which translation is constitutively inhibited.

Select constructs were tested to determine whether accessibility of the SD region corresponds to the efficiency of translational repression observed during growth in the presence and absence of thiamin using an oligonucleotide-directed RNase H protection assay. Also, mutations that affected thiamin-dependent repression in vivo were tested in primer extension inhibition assays to monitor 30S ribosomal subunit binding in the presence and absence of TPP. We show that TPP binding to the wild-type thiM leader RNA results in a conformational change that causes occlusion of the SD sequence, and inhibition of ribosome binding, consistent with the model. Alteration of conserved residues in the ligand binding domain affected accessibility of the SD sequence and binding of 30S ribosomal subunits (in the presence and absence of TPP), consistent with the effects on expression that were observed in vivo. These results directly demonstrate the molecular basis for THI box gene regulation, and confirm the roles of several conserved residues in ligand recognition and regulatory control.
2.2 Materials and Methods

2.2.1 DNA constructs

DNA templates for T7 RNA polymerase (RNAP) transcription reactions, corresponding to residues 11-168 and 11-239 of the *E. coli thiM* leader sequence, were constructed by annealing and ligating overlapping pairs of complementary oligonucleotides (Integrated DNA Technologies, Coralville, IA) as described previously (McDaniel et al., 2005; Yousef et al., 2003). Oligonucleotide sequences are shown in Table 2.1. Oligos used to generate the 5’ end of the *E. coli thiM* leader RNA (EcThiM casA and EcThiM casB) contained a sequence corresponding to a T7 RNAP promoter followed by an additional G residue to facilitate efficient transcription initiation.

Complementary oligonucleotides containing nucleotide substitutions at desired residues were used to generate mutant *E. coli thiM* constructs. Templates were amplified by PCR, purified using a QIAquick PCR purification kit (Qiagen), and verified by DNA sequencing.
2.2.2 RNase H probing of the SD region of *E. coli thiM* transcripts

Radiolabeled RNAs corresponding to positions 11–168 of the *E. coli thiM* transcript were synthesized using a MEGA-script T7 RNAP transcription kit (Ambion) in the presence of [α-32P]-UTP. RNAs generated in the presence or absence of TPP (500 µM) were incubated with DNA oligonucleotide EcThiM RNH6 (Table 2.1; 45 nM), complementary to positions 125–133, for 5 min at 37 °C followed by addition of RNase H (0.45 U ml⁻¹; Ambion). Digestion was terminated after 10 min at 37 °C by addition of phenol and cleavage products were resolved on a 6% denaturing polyacrylamide gel. Digestion products were visualized by PhosphorImager analysis (Molecular Dynamics) and quantified using ImageQuant 5.2 software by determining the percentage of cleavage product relative to the total amount of both cleaved and full-length RNA. The percent
cleavage for each mutant RNA was normalized to that observed for the wild-type *E. coli* *thiM* transcript in the absence of TPP, and all experiments were carried out in triplicate.

### 2.2.3 Toeprint assay of *E. coli thiM* transcripts with 30S ribosomal subunits

DNA templates for T7 RNAP transcription were prepared as described above. RNAs corresponding to positions 11–239 of the *E. coli thiM* transcript were synthesized using an AmpliScribe T7 High Yield Transcription Kit (Epicentre Biotechnologies). The reactions were treated with DNase and the RNAs were purified on a 6% denaturing polyacrylamide gel and electroeluted in RNase-free TE buffer. DNA oligonucleotide EcThiM RT1 (complementary to positions 211–234 of the *E. coli thiM* transcript, see Table 2.1) was 5′-end labeled with [γ-32P]-ATP (7000 Ci mmol⁻¹) using a KinaseMax kit (Ambion) and passed through a G25 MicroSpin column (Amersham Biosciences) to remove unincorporated [γ-32P]-ATP. For toeprinting assays, end-labeled oligonucleotide (10 nM) was annealed to *E. coli thiM* transcript (10 nM) in 1X binding buffer (10 mM Tris-Cl, pH 7.5, 10 mM MgCl₂, 60 mM NH₄Cl and 6 mM 2-mercaptoethanol) by heating to 65 °C for 5 min followed by slow cooling to 40 °C. TPP (500 µM) or RNase-free ddH₂O was added to the annealed RNA-oligonucleotide complex and the mixture was incubated at 37 °C for 10 min followed by addition of initiator tRNAfmet (900 nM; Sigma) and/or 30S ribosomal subunits (450 nM) isolated from *E. coli* strain MRE600. Ribosomal subunits were prepared with a high salt wash as previously described (Powers and Noller, 1991), and were provided by Dr. Kurt Fredrick (The Ohio State University). After 30 min incubation at 37 °C, dNTPs (40 nM) and ThermoScript Reverse Transcriptase (0.05 U ml⁻¹) were added to each reaction to initiate cDNA synthesis. Reverse transcription was
terminated after an additional 10 min incubation at 37 °C by addition of gel loading buffer (Ambion) and the cDNA products were resolved on a 10% denaturing polyacrylamide gel. A DNA sequencing ladder was generated as a size standard by using a Sequenase 2.0 Kit (USB) with the same DNA template used for RNA transcription and oligonucleotide EcThiM RT1 as primer. All products were visualized using PhosphorImager analysis (Molecular Dynamics).

2.2.4 TPP binding assays

RNAs corresponding to positions 11–168 of the E. coli thiM transcript were synthesized by T7 RNAP transcription and gel purified as described above. RNAs (8 mM) were suspended in 1X transcription buffer (20 mM Tris-Cl, pH 8.0, 20 mM NaCl, 10 mM MgCl₂, 0.1 mM EDTA), heated to 65 °C and slow cooled to 40 °C. TPP (8 mM) was added to the RNA, and reactions were incubated at 37 °C for 30 min. The binding mixtures were passed through a Nanosep 10K Omega filter and the concentration of unbound TPP in the flowthrough fraction was compared with the concentration of TPP in the original mixture after oxidization to a fluorescent derivative using potassium ferricyanide (8.6 mM K₃[Fe(CN)₆] in 1.4 M NaOH; Kawasaki, 1986; Batifoulier et al., 2005). A standard curve was generated using known amounts of TPP, and a linear response was observed for samples containing 10-500 pmol. Fluorescence intensity at 455 nm was measured on a FlexStation3 Microplate Reader (Molecular Devices) with an excitation wavelength of 365 nm. The percentage of bound TPP was determined relative to the total amount of TPP for each reaction.
2.3 Results

2.3.1 In vivo expression analysis of wild-type and mutant THI box RNAs.

*In vivo* expression experiments were performed in the laboratory of Dr. Juan Miranda-Rios (Universidad Nacional Autónoma de México). To investigate the effects of mutations at conserved residues in the THI box ligand binding domain on TPP-dependent repression *in vivo*, our collaborators utilized an *E. coli thiM-lacZ* translational fusion in which β-galactosidase expression is dependent on translation initiation occurring at the *thiM* ribosome binding site (Fig. 2.3A). The *E. coli thiM* leader and variant constructs were cloned into plasmid pRS552 (Simons et al., 1987) such that the 6th codon of *thiM* was fused in frame with the 7th codon of a *lacZ* reporter. Recombinant plasmids were transformed into *E. coli* and the level of *lacZ* expression was monitored by measuring the β–galactosidase activity from cells grown in the presence or absence of thiamin. In this system, thiamin provided in the media is converted to TPP through the thiamin biosynthesis pathway. Highly conserved residues located at the interior loops within the THI box ligand binding domain were chosen for mutagenesis to investigate the effects of these mutations on TPP-dependent repression *in vivo* (Fig. 2.3B). Each of the selected residues was found to be at least 97% conserved among 218 THI box motifs identified in 106 prokaryotic genomes (Abreu-Goodger et al., 2004). In addition to these highly conserved residues, U79 was selected for comparison since it is conserved in only 56% of the known THI box riboswitches.
Fig. 2.3. *In vivo* expression analysis of wild-type and mutant THI box RNAs. A) Plasmid-borne translational fusions containing the wild-type or mutant *thiM* leader sequences fused to a *lacZ* reporter were introduced into *E. coli* by electroporation. The effect of TPP on *thiM-lacZ* expression was monitored by growing cells in the presence or absence of thiamin and testing for β-galactosidase activity. B) Conserved residues that were mutated and tested in translational fusions are boxed. Residues whose mutation yielded a Class I repression phenotype are indicated in green, Class II in red, and the U79 residue is shown in orange. C) β-galactosidase activity is expressed in Miller units (Miller, 1972). Values represent the averages from three independent experiments (mean ± standard error). Colored bars correspond to phenotypic classes as in part (B). V, plasmid pRS552 without insert. *In vivo* expression analysis experiments were performed by the lab of Dr. Juan Miranda-Rios, Universidad Nacional Autónoma de México.
Expression of the wild-type *thiM-lacZ* fusion was repressed by more than 20-fold in cells grown in the presence of thiamin (Fig. 2.3C). Mutation at each of the conserved residues in the ligand binding pocket resulted in a loss of TPP-dependent repression *in vivo*. The majority of the mutant constructs, referred to as Class I (G19C, C38G, U39A, G40C, A41U, G42C, A43U, G60C, and U68A), exhibited high expression in both the presence and absence of thiamin. This phenotype is consistent with a loss of TPP binding, or of a loss of the structural rearrangement of the RNA that is required for sequestration of the SD sequence. In contrast, Class II constructs (A61U, C77G, and G78C) exhibited low expression even in the absence of TPP, suggesting that these RNAs favor a conformation in which the SD sequence is always sequestered. Finally, the U79A mutation exhibited a somewhat intermediate phenotype with decreased expression of β-galactosidase activity relative to wild-type, and more than two-fold repression in response to thiamin addition to the growth media. These data revealed that mutation of residues in the ligand binding pocket affects the regulatory output of the THI box riboswitch *in vivo*. However, it was unclear whether these effects were a result of changes in TPP binding affinity, or in the structural rearrangement capabilities of the RNA.

2.3.2 TPP decreases sensitivity of the *E. coli* thiM SD region to oligonucleotide-directed RNase H cleavage

RNase H, an enzyme that specifically cleaves RNA-DNA hybrids, was used to investigate the accessibility of the SD sequence of the *E. coli* thiM transcript. An oligonucleotide complementary to the SD region was added to a radiolabeled transcript (corresponding to positions 11–168) that was synthesized by T7 RNAP in the presence or
absence of TPP. RNase H was added to cleave RNA annealed to the DNA oligonucleotide and digestion products were resolved using denaturing PAGE. In the absence of TPP, the wild-type *E. coli thiM* RNA was efficiently cleaved by RNase H (Fig. 2.4A and B), indicating that the SD region is accessible to binding of the DNA oligonucleotide. Oligonucleotide-directed RNase H cleavage was reduced approximately 3-fold when the RNA was transcribed in the presence of TPP, suggesting that addition of TPP causes a structural change that results in decreased accessibility of the SD region of the RNA to the DNA oligonucleotide, consistent with the model and with previous studies.
Fig. 2.4. Oligonucleotide-directed RNase H cleavage of E. coli thiM RNA. A) Denaturing PAGE analysis of RNase H cleavage products. Oligonucleotide EcThiM RNH6, complementary to residues 125–133 (which includes the SD sequence) was added to radiolabeled E. coli thiM RNA generated in the presence (+) or absence (−) of TPP (500 µM). RNA-DNA hybrids were cleaved with RNase H and the products were visualized by autoradiography. B) Quantification of RNase H cleavage products relative to that observed for wild-type RNA in the absence of TPP. The values plotted represent the means ± standard deviations of at least three independent experiments for each construct. FL, full-length transcripts; C, RNase H cleavage products.

We next selected two constructs pertaining to Class I and Class II with mutations in the TPP binding pocket. Mutant constructs with single nucleotide changes in the transcript sequence were tested using the RNase H assay to determine whether accessibility of the SD region corresponds to efficiency of translational repression during growth in the presence of thiamin. A thiM transcript containing the G40C mutation,
which resulted in a loss of repression by thiamin in vivo, exhibited efficient oligonucleotide-directed RNase H cleavage regardless of the presence of TPP (Fig. 2.4). Similar results were observed with a thiM variant containing the G60C mutation, which also showed loss of thiamin-dependent repression in vivo. These results indicate that in the G40C and G60C mutant transcripts, the SD region is accessible regardless of the presence of TPP, consistent with constitutive high expression in vivo. In contrast, the G78C variant that exhibited low expression in vivo even under low thiamin conditions, demonstrated decreased RNase H cleavage in the presence or absence of TPP. This suggests that in this variant the SD sequence is inaccessible to the antisense oligonucleotide regardless of the presence of TPP. The A61U mutant, which also exhibited constitutive repression in vivo, demonstrated decreased RNase H cleavage in the absence of TPP, but to a lesser degree than the G78C mutant. The decreased accessibility of the SD sequence observed in the G78C and A61U variants relative to the wild-type thiM RNA is consistent with the low expression observed in vivo in the presence or absence of thiamin.

2.3.3 30S ribosomal subunit binding to the E. coli thiM transcript

A primer extension inhibition (‘toeprint’) assay was used to investigate the effect of TPP on binding of the 30S ribosomal subunit to the E. coli thiM ribosome binding site. A radiolabeled DNA oligonucleotide complementary to residues 211–234 (downstream of the SD) was annealed to an RNA corresponding to positions 11–239 of the thiM sequence, followed by addition of tRNA^{Met} and 30S ribosomal subunits. Reverse transcriptase and dNTPs were then added to initiate reverse transcription. In this
method, cDNA synthesis proceeds until the enzyme encounters an obstacle such as a ribosome bound to the transcript, or significant structure within the RNA template. cDNA products were resolved by denaturing PAGE in parallel with a DNA sequencing ladder used as a size standard.

The full-length product, which corresponds to cDNA extending to the 5′ end of the \textit{thiM} transcript, was observed in all reactions (Fig. 2.5). A second product corresponding to a stop at position A131 relative to the transcription start-site (two nucleotides downstream from the SD) was also observed in all reactions. Addition of TPP to the wild-type \textit{thiM} RNA resulted in an increase in the abundance of extension products corresponding to position A131, suggesting that RNA structure in this region is enhanced when TPP is added. The most dramatic effect of TPP addition was the appearance of an abundant extension product at position G86 (Fig. 2.5, lanes 2 and 6). TPP-dependent inhibition of extension at this position is not surprising since G86 is located within the P1 helix located at the base of the ligand binding pocket, which is predicted to become stabilized upon TPP addition (Lang et al., 2007; Rentmeister et al., 2007).
Fig. 2.5. Primer extension inhibition analysis of *E. coli thiM* RNA. Transcripts generated by T7 RNAP transcription were incubated with radiolabeled oligonucleotide EcThiM RT1 (complementary to residues 211–234) followed by addition of TPP as indicated. tRNA<sup>Met</sup> and 30S ribosomal subunits were added as indicated, and extension reactions were initiated by addition of dNTPs and reverse transcriptase. Reactions were halted by addition of gel loading buffer and products were resolved by PAGE and visualized by autoradiography. Prominent extension products were identified using a DNA sequencing ladder (far right) as a size standard and are indicated on the left. This figure was assembled from several separate gels. FL, full-length products.
Addition of 30S ribosomes and tRNA$_{fmet}$ to wild-type *E. coli thiM* RNA resulted in a new extension product corresponding to position C151 (Fig. 2.5, lane 5). This product was not observed in reactions performed in the absence of either tRNA$_{fmet}$ or 30S ribosomal subunits. Position C151 is 16 nt downstream of the AUG start codon, which represents the location of a predicted reverse transcriptase stop when the start codon is positioned within the P site of the ribosome. The abundance of extension products at this position decreased in the presence of TPP, supporting the model that TPP prevents binding of the 30S ribosomal subunit to the RNA. These data are in agreement with the RNase H probing experiments that demonstrated that the SD region is less accessible in the presence of TPP.

*E. coli thiM* constructs containing mutations that affected thiamin-dependent repression *in vivo* were tested in primer extension inhibition assays to investigate 30S ribosomal subunit binding in the presence and absence of TPP. The G40C and G60C constructs, which exhibited a loss of repression during growth in high thiamin, were able to bind the ribosome with equal affinity in the presence or absence of TPP, as indicated by the abundance of the extension product at position C151 (Fig. 2.5, lanes 7–10). This is in agreement with the results obtained from the RNase H assay that showed that the SD region of these two transcripts was accessible even when TPP was added. The abundance of extension products corresponding to position G86 decreased significantly relative to that observed for the wild-type RNA, suggesting that structural features stabilizing the P1 helix at the base of the metabolite binding pocket were disrupted in these mutants. These data are consistent with crystal structures of the *E. coli thiM* RNA-TPP complex that showed that both G40 and G60 are involved in TPP binding (Edwards and Ferre-
D'Amare, 2006; Serganov et al., 2006). The extension product at position A131 was present in all reactions containing the G40C and G60C mutants and, similar to reactions performed with wild-type RNA, this product increased in abundance upon addition of TPP. This suggests that some modulation of RNA structure in this region occurs in the presence of TPP with these mutant constructs. Reactions containing the G40C mutant RNA resulted in an additional extension product corresponding to position G51 (Fig. 2.5, lanes 7, 8), which lies at the base of the P3 helix in the TPP binding domain. This extension product occurred regardless of the presence of TPP and indicates that structural features in this region of the RNA may be altered relative to wild-type.

The G78C and A61U mutant constructs, which exhibited low expression in vivo independent of the addition of thiamin, were also tested. The G78C mutant showed a decrease in ribosome binding relative to the wild-type RNA, regardless of the presence of TPP, as indicated by the low abundance of extension products corresponding to position C151 (Fig. 2.5, lanes 11, 12). This is consistent with the RNase H data which showed that the SD region of this construct was protected from cleavage in the presence or absence of TPP. Additionally, the TPP-dependent extension product corresponding to position G86 was not observed, suggesting that structural features within the RNA P1 helix are altered. Reactions containing the A61U mutant RNA resulted in a small amount of the G86 extension product and a decrease in the 30S ribosomal subunit toeprint upon addition of TPP (Fig. 2.5, lanes 13, 14). These results indicate that the SD region of the A61U mutant is partially accessible in the absence of TPP, and that this construct retains some ability to form the P1 helix when TPP is added. Although the A61U transcript is able to bind the 30S ribosomal subunit under the reaction conditions tested here (in the absence
of TPP), this interaction appears to be insufficient to support efficient translation initiation as evidenced by low gene expression \textit{in vivo}, and partial occlusion of the SD sequence in the RNase H cleavage assay. For both mutants, the decrease in the abundance of the G86 extension product relative to wild-type suggests that the TPP-induced stabilization of the P1 helix at the base of the ligand binding domain is disrupted. The results obtained with the G78C construct suggest that this mutation results in an RNA structure in which the SD region is sequestered, leaving the ribosome binding site inaccessible to the 30S ribosomal subunit regardless of the presence of TPP.

2.3.4 Binding of TPP by the \textit{E. coli} thiM transcript

We developed a binding assay to determine whether mutation of residues in the ligand binding domain of the THI box RNA disrupts binding to TPP. \textit{thiM} leader RNAs generated by T7 RNAP transcription were incubated with TPP, and unbound TPP was removed by filtration and quantified relative to the input amount of TPP. TPP detection was accomplished using a fluorescence method in which TPP is oxidized to a thiochrome derivative in the presence of potassium ferricyanide (Batifoulier et al., 2005; Kawasaki, 1986). The wild-type \textit{thiM} leader RNA bound approximately 40% of the input TPP in three independent experiments (normalized to 100% in Fig. 2.6). Introduction of the G40C substitution resulted in a complete loss of TPP binding. Residue G40, which in the wild-type structure directly contacts the pyrimidine moiety of TPP through two hydrogen bonds (Edwards and Ferre-D'Amare, 2006; Serganov et al., 2006), is therefore essential for TPP binding. Another Class I mutation, G60C, resulted in a 3-fold reduction in TPP binding relative to the wild-type RNA. This confirms that G60, which was shown in the
crystal structure to interact with TPP indirectly through a divalent cation, is also important for TPP binding. Disruption of TPP binding in the G40C and G60C mutant constructs is consistent with the loss of the TPP-dependent conformational change required for masking of the SD sequence. The Class II mutant G78C demonstrated a modest decrease (less than two-fold) in TPP binding relative to the wild-type RNA (Fig. 2.6). The other Class II mutant construct tested, A61U, exhibited TPP binding similar to wild-type RNA. Although both of these mutant constructs retained the ability to bind TPP, ligand-induced conformational changes occurring at the SD-ASD region and repression of gene expression were altered relative to wild-type. These data support the model that in the Class II mutants the RNA favors a conformation in which the SD-ASD pairing is stabilized independent of the presence of TPP.

Fig. 2.6. TPP binding to E. coli thiM transcripts. RNAs generated by T7 RNAP transcription were denatured, refolded and incubated with TPP. Unbound TPP was separated from bound TPP by filtration and was quantified relative to the amount of input TPP. TPP was oxidized to a thiochrome derivative using potassium ferricyanide and fluorescence was measured at 455 nm with an excitation wavelength of 365 nm. Percent binding for each construct represents average TPP binding in at least three independent experiments and is expressed relative to that observed with the wild-type thiM leader RNA.
2.4 Discussion

Recently, crystal structures of the *A. thaliana* thiC and *E. coli* thiM riboswitches were published, providing crucial insight into ligand-RNA interactions and tertiary folding of the TPP binding domain. However, these studies were done using limited constructs lacking downstream residues that constitute the expression platform where the regulatory event occurs. While several studies have focused on the structural rearrangements that the riboswitch undergoes in response to TPP binding (Lang et al., 2007; Mayer et al., 2007; Rentmeister et al., 2007), limited evidence has been presented verifying the mechanism of regulation by the THI box riboswitch, and exploring the molecular interplay between riboswitch domains. Additionally, while the crystal structures revealed structural elements important for ligand recognition, a thorough mutational analysis has not been reported previously.

In these studies, we probed the accessibility of the SD sequence of the wild-type *E. coli* thiM leader sequence, in the presence and absence of TPP, using an oligonucleotide-directed RNase H protection assay. These studies revealed that the SD sequence exhibits decreased accessibility when TPP levels are high, supporting the model that translation initiation is diminished in response to ligand binding. In addition, we directly demonstrated for the first time, using a primer extension assay, that 30S ribosomal subunit binding to the THI box element is decreased in the presence of TPP. Our data verify the location of the thiM ribosome binding site, as shown by the appearance of a “toeprint” at residue C151 (which is 16 nt downstream of the predicted AUG start codon) when 30S ribosomal subunits and tRNA^fmet_ were added to the primer extension reaction.
Our collaborators at the Universidad Nacional Autónoma de México identified conserved bases located in the ligand binding pocket of the *E. coli thiM* riboswitch, and investigated the effect of mutation at those residues on expression of a *thiM-lacZ* translational fusion, during growth in the presence and absence of thiamin. As expected, expression of the wild-type *thiM-lacZ* fusion was repressed in the presence of thiamin. Mutation at any of the selected residues resulted in a loss of TPP-dependent repression, and two distinct phenotypes were observed. The majority of the mutant constructs exhibited a Class I phenotype in which gene expression was constitutively high regardless of the presence or absence of thiamin. Loss of repression in these mutant RNAs might be explained by a loss of TPP binding affinity. Conversely, loss of repression might be due to the inability of the RNA to undergo the structural rearrangement required for sequestration of the SD sequence. In order to test these two possibilities, we selected two representative Class I mutant RNAs (G40C and G60C; Fig. 2.7), and tested them for their ability to bind TPP and undergo conformational changes at the SD sequence in response to ligand binding.
Fig. 2.7. Proximity of selected residues in the THI box ligand binding pocket. Four residues were selected for further analysis to determine how changes at these positions affect TPP binding and accessibility of the SD region. Alteration of G40 and G60 (green) resulted in a Class I phenotype (high constitutive expression \textit{in vivo}), and alteration of A61 and G78 (red) resulted in a Class II phenotype (low uninducible expression \textit{in vivo}). Residues are shown relative to TPP (center) and water coordinated divalent cations (green balls).

The G40 residue makes direct contact with the pyrimidine moiety of TPP, while the G60 residue interacts with a divalent cation that binds to the pyrophosphate group (Fig. 2.7). Because of their roles in ligand recognition it was predicted that alteration of these residues would result in a decrease in TPP binding. Indeed, the two Class I mutant constructs G40C and G60C exhibited a significant decrease in TPP binding as compared to the wild-type RNA. Consistent with these findings, addition of TPP failed to result in occlusion of the SD sequence (as determined by RNase H cleavage experiments), or inhibition of 30S ribosomal subunit binding to the mRNA (as determined by toeprinting assays). Taken together, these data suggest that the loss of TPP-dependent repression in these two Class I mutants is likely to be due to loss of TPP binding. Without the ability to
bind TPP, these RNAs fail to undergo the structural rearrangement required for sequestration of the SD sequence, and gene expression remains high even during growth under conditions of high thiamin.

In addition to the G40C construct, three other Class I mutants (G19C, G42C, and A43U) contain nucleotide substitutions at residues that directly contact the pyrimidine moiety of TPP (Edwards and Ferre-D'Amare, 2006; Serganov et al., 2006). In contrast, the C38G, U39A, and U68A constructs, which also exhibit a Class I phenotype, contain substitutions at residues that are predicted to be important for stabilizing a tertiary interaction between helix P3 and loop L5 of the riboswitch RNA when TPP is bound. This tertiary interaction is predicted to stabilize compaction of the riboswitch RNA prior to formation of helix P1. Lastly, the A41U construct, which contains a mutation that is predicted to disrupt a base triple interaction with the G18-C48 base pair at the top of helix P2 (Serganov et al., 2006), also exhibits a Class I phenotype. As with the G40C and G60C mutants, a decrease in TPP binding affinity might explain the loss of repression in these other Class I constructs. However, additional experiments are needed in each case to determine whether the observed phenotype is a result of decreased ligand binding or a loss of the RNA structural rearrangement required for formation of the P1 helix.

Several constructs, referred to as Class II, exhibited an alternate phenotype in which gene expression was repressed during growth in both the presence and absence of thiamin. These results suggest that these RNAs are unable to assume the unliganded conformation even in the absence of TPP, and that the SD sequence is always occluded. To verify this hypothesis, we selected two representative Class II constructs (A61U and G78C; Fig. 2.7) to investigate the accessibility of their SD sequences in the presence and
absence of TPP, and to test their ability to bind TPP. G78 is the only residue in the THI box RNA to make direct contact with the TPP pyrophosphate moiety, while A61 interacts indirectly, through a water molecule, with a divalent cation that binds TPP (Edwards and Ferre-D'Amare, 2006). Despite their roles in ligand recognition, substitution of these residues in the G78C and A61U constructs did not significantly affect ligand binding. However, while these constructs retain the ability to bind TPP, their SD sequences exhibited decreased accessibility (relative to the wild-type RNA) in the absence of TPP. Also, in the case of the G78C construct, 30S ribosomal subunit binding was inhibited as indicated by a decrease in the abundance of primer extension products at position C151 even in reactions lacking TPP. These results verify that the substitutions at residues A61 and G78 are sufficient to cause the RNA to favor a conformation in which the SD sequence is sequestered regardless of the presence of the ligand. A similar THI box mutation that results in repression of gene expression even in the absence of TPP was previously reported (Rentmeister et al., 2007). However, in that instance, the nucleotide substitution was located in the THI box expression platform and not in the TPP binding domain. Therefore, the Class II mutations described here represent a novel type of riboswitch mutation in which substitutions in the ligand binding domain do not affect ligand binding, but instead cause the RNA to favor a conformation mimicking the ligand-bound state even in the absence of ligand. Additional studies are needed to verify that the third Class II construct, G77C, also exhibits decreased accessibility of the SD sequence, as predicted based on in vivo results. Also, it is not known whether or not this construct retains the ability to bind TPP.
The P2/P3 bulge of the THI box ligand binding domain has been shown to be preformed in the absence of TPP (Rentmeister et al., 2007). The P4 and P5 helices are not formed in the absence of ligand, and instead an alternate helix is formed that includes pairing between the AASD and the ASD regions, leaving the SD sequence free to bind the ribosome (see Fig. 2.1). Addition of TPP results in binding of the pyrimidine moiety to the preformed P2/P3 bulge, and stabilization of helices P4 and P5. TPP contacts the P4/P5 bulge via its pyrophosphate moiety, and tertiary interactions between loop L5 and helix P3 help to compact the ligand binding domain. These latter events are predicted to lead to stabilization of helix P1 at the base of the ligand binding pocket, which sequesters the AASD sequence thereby freeing up the ASD sequence for sequestration of the SD sequence (Lang et al., 2007; Rentmeister et al., 2007). The vast majority of Class I mutations identified in this study reside in the P2/P3 bulge. Many of these residues either make direct contacts with the TPP pyrimidine moiety, or form tertiary interactions that help tether the parallel helices that encompass the TPP molecule. Mutation of these residues results in a loss of TPP-dependent repression \textit{in vivo}, and in many instances (as in the case of the G40C and G60C constructs), this is likely to be due to a loss of ligand binding, which is required for conferring the regulatory response to the expression platform downstream. In contrast, all of the Class II mutations identified here involve nucleotide substitutions at residues in the P4/P5 bulge that either directly or indirectly contact TPP through its pyrophosphate moiety. While nucleotide substitutions at G78 and A61 did not alter TPP binding, mutation at these residues resulted in constitutive repression of gene expression \textit{in vivo}. This indicates that this second type of mutation results in a tweaking of the riboswitch structure such that the SD sequence is sequestered
even in the absence of ligand. Taken together, these data show that mutation of conserved residues important for the recognition of the pyrimidine group of TPP often results in a loss of ligand binding, whereas mutation of residues involved in the recognition of the pyrophosphate moiety are more tolerated for ligand binding, but can dramatically alter the conformational state of the riboswitch RNA.
CHAPTER 3

CRYSTAL STRUCTURE OF THE S\textsubscript{MK} BOX RIBOSWITCH IN COMPLEX WITH LIGAND

3.1 Introduction

Three known riboswitch classes (S box, SAM-II, and S\textsubscript{MK} box) that bind the enzymatic cofactor SAM have been identified (Corbino et al., 2005; Fuchs et al., 2006; McDaniel et al., 2003). SAM is comprised of an adenosine group, a positively charged sulfonium moiety, and a methionine tail (Fig. 3.1). Attached to the positively charged sulfonium is a chemically reactive methyl group that acts as a donor in many important methyltransferase reactions inside the cell. Each of the SAM-responsive riboswitches is able to selectively recognize SAM and discriminate against the closely related metabolic byproduct SAH, which lacks the methyl group and positive charge on the sulfonium moiety (Fig. 3.1A). The crystal structures of the S box and SAM-II riboswitches verified that these two RNAs have ligand binding pockets that are completely different from one another and utilize distinct mechanisms for ligand recognition (Gilbert et al., 2008; Montange and Batey, 2006). Although no crystal structure has previously been reported for the S\textsubscript{MK} box riboswitch, this RNA is also predicted to utilize a unique architecture for SAM recognition based on phylogenetic and mutational analyses (Fuchs et al., 2006).
Determination of the high resolution structure of this third SAM binding riboswitch will provide a comprehensive picture for the comparison of these three RNAs that have evolved independently to recognize the same molecular signal.

The SMK box riboswitch has been identified in the 5’ UTR of the metK gene in many lactic acid bacteria (Fuchs et al., 2006, 2007). metK encodes SAM synthetase, the enzyme responsible for the formation of SAM from methionine and ATP. All SMK box elements identified to date are predicted to regulate at the level of translation initiation by occlusion of the TIR in response to SAM binding (Fig. 3.1B). Specifically, the last three nucleotides of the SD sequence along with two additional residues are predicted to be sequestered by an upstream ASD sequence when SAM is bound. In the absence of ligand, the RNA is predicted to favor an alternate conformation in which both the ASD and SD sequences are unpaired. In this arrangement, the SD sequence is available for binding to the ribosome and translation initiation can occur.
Fig. 3.1. Chemical structures of SAM and SAH and predicted secondary structure of the *E. faecalis* S_MK box riboswitch. A) SAH differs from SAM by a single methyl group and positive charge on the sulfonium moiety (highlighted in pink). B) Model of the riboswitch secondary structure based on phylogenetic and mutational analyses. The S_MK box riboswitch from *E. faecalis* regulates at the level of translation initiation. In the absence of SAM (left) the RNA assumes a conformation in which the SD sequence (black bar) is accessible for translation initiation. When SAM is bound (right) the RNA favors a conformation in which the ASD sequence (red) base pairs with a sequence that overlaps the SD sequence (green) and expression of *metK* is repressed. The AUG start codon is indicated in yellow. Adapted from (Fuchs et al., 2006).
The *metK* leader RNA from *E. faecalis* has been the focus of several key investigations to characterize the regulatory mechanism employed by this riboswitch. SAM-dependent repression of an *E. faecalis metK-lacZ* translational fusion integrated into the *B. subtilis* chromosome provided strong evidence that regulation conferred by this riboswitch is likely to occur at the level of translation (Fuchs et al., 2006). Direct *in vitro* evidence for the translational mechanism of regulation was obtained using nitrocellulose filter binding and ribosomal toeprinting assays, which showed that 30S ribosomal subunit binding is inhibited in the presence of SAM (Fuchs et al., 2007).

Two possible alternate conformations were proposed for the *E. faecalis* SMK box riboswitch (Fig. 3.1B, Fuchs et al., 2006). Areas of potential pairing in addition to the complementarity observed between the SD and ASD regions were identified based on sequence alignments. Results from RNase H cleavage experiments verified that the SD and ASD sequences are occluded in the presence of SAM (Fuchs et al., 2006). Decreased accessibility was also observed around residues 29-32 and 67-69 (located in the core of the molecule) in the presence of SAM, suggesting that these residues form base pairs consistent with the model. Results from structural mapping experiments provided additional evidence for key secondary structural features of the *E. faecalis* SMK box element.

In addition to defining the secondary structure of the SMK box riboswitch our lab has also been interested in defining the sequence determinants for SAM binding. A construct corresponding to residues 15-118 relative to the predicted transcription start site exhibits efficient SAM binding *in vitro* and has been used for comparison to investigate the effect of mutations on ligand recognition. Truncated transcripts corresponding to
residues 15-95 and 20-118 retain the ability to bind SAM (Fuchs et al., 2006, 2007). However, additional truncations at the 5’ and 3’ ends, which remove residues within the SD and ASD sequences, result in a complete loss of SAM binding indicating that these residues are essential for ligand recognition (Fuchs et al., 2006). Additional deletion constructs were tested to determine the importance of residues residing in the long hypervariable regions (i.e., residues in the upper part of the core element and just upstream of the SD sequence) for SAM binding. Complete deletion of the terminal stem loop (residues 34-61) at the top of the SMK box element resulted in a 10-fold reduction of SAM binding, while shortening of the stem loop (deletion of residues 36-44 and 51-59) resulted in only a modest decrease in SAM binding (82% binding activity relative to wild-type; R.T. Fuchs, unpublished). Deletion of residues 77-86 upstream of the SD sequence also resulted in only a modest decrease in SAM binding (75% binding activity relative to wild-type; R.T. Fuchs, unpublished), indicating that residues in this region are not essential for ligand recognition.

Mutational analyses revealed that base pairing between the SD and ASD sequences is required for SAM binding to the E. faecalis metK leader (Fuchs et al., 2006, 2007), suggesting that the regulatory domain is intrinsic to the ligand binding domain. The only other riboswitch that is predicted to exhibit a similar single-domain architecture is the SAM-II riboswitch, although the regulatory mechanism of this RNA has not been confirmed. In contrast, the S box riboswitch exhibits a modular architecture (similar to all other known riboswitch classes) that involves the interplay between two discrete domains, one for ligand binding and the other for regulatory output.
Using information garnered from previous mutational analyses, we designed and tested a variety of deletion constructs to define a minimal $S_{MK}$ box RNA that retains the ability to bind SAM. A 53-nt construct was used to determine the 2.2-Å crystal structure of the $S_{MK}$ box riboswitch in complex with SAM (Lu et al., 2008). Crystallography experiments were performed by the laboratory of Dr. Ailong Ke (Cornell University, Ithaca, NY). The crystal structure confirmed key structural features of the riboswitch, and showed that SAM makes direct contacts with residues in the SD sequence and other conserved nucleotides in the ligand binding pocket. We performed mutational analyses to verify RNA sequence determinants required for SAM recognition (done by R.T. Fuchs and A.M. Smith).

3.2 Materials and Methods

3.2.1 RNA synthesis and purification

Oligonucleotide sequences used for full-length and truncated constructs are shown in Table 3.1. DNA templates for T7 RNAP transcription reactions, corresponding to residues 15-118 of the *E. faecalis metK* leader sequence, were constructed by annealing and ligating overlapping pairs of complementary oligonucleotides (names starting with EfMetK cas in Table 3.1; Integrated DNA Technologies, Coralville, IA) as described previously (McDaniel et al., 2005; Yousef et al., 2003). Oligos used to generate the 5’ end of the full-length *metK* leader contained a T7 RNAP promoter sequence followed by an additional G residue to facilitate efficient transcription initiation. Complementary oligonucleotides containing nucleotide substitutions at desired residues (not shown) were used to generate full-length mutant constructs. Templates were
amplified by PCR, purified using a QIAquick PCR purification kit (Qiagen), and verified by DNA sequencing.

<table>
<thead>
<tr>
<th>Primer</th>
<th>Sequence (5’ to 3’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>EfMetK casA</td>
<td>TAATACGACTCACTATAGGGTACACGTAGTCAAGGAGTCCCGAAGGATTTAAGCAAG</td>
</tr>
<tr>
<td>EfMetK casB</td>
<td>TAAATCTCTTTCGGTGAACCTTGTAATACGTGAGTCGATTTA</td>
</tr>
<tr>
<td>EfMetK casC</td>
<td>TAATGGCTCATTACTAAAGATGCGCTTGGTAACGAG</td>
</tr>
<tr>
<td>EfMetK casD</td>
<td>TTACAAAGGACATTTAGTAGAACGACAAATTACGTC</td>
</tr>
<tr>
<td>EfMetK casE</td>
<td>AAGTTATCATGCGCAATTTAGTGATGTACGGCATCTGAC</td>
</tr>
<tr>
<td>EfMetK casF</td>
<td>AAATGCTTTCCTGCTCATATGGTATTTATCCCCTAATAGTTCCG</td>
</tr>
<tr>
<td>MetKshort/RZ 5-1</td>
<td>AACGCGCAGATGCGTCTTGATAACCCAAAAGAGGAGGGCCGATCGCTCCC</td>
</tr>
<tr>
<td>MetKshort/RZ 5-2</td>
<td>GTACCTCGGTACCGCTTCCGGAAAGGATGCGGGAACGCGGAGCCAGGATCGCTTGG</td>
</tr>
<tr>
<td>MetKshort/RZ 5-3</td>
<td>TCGGGAACATGATGTCGCGTGAGCGCAAGCCGAGGCTACCGTACGGTTCC</td>
</tr>
<tr>
<td>MetKshort/RZ 5-4</td>
<td>GGCTGAAATACGACGAAATGCTATAGCCTTTCGAGGTGGAATACACGAACAAAG</td>
</tr>
<tr>
<td>MetKshort/RZ 5-5</td>
<td>GCCATCGATCTTCAATACGCTACATAGGGGTGATACACGCAACAAAG</td>
</tr>
<tr>
<td>MetKshort8A/RZ 3-1</td>
<td>GGCTGGGACCATGCGCGCCCTTCCCTTCTTTTGTTGTTACAGGCGATCTTGCA</td>
</tr>
<tr>
<td>MetKshort7A/RZ 3-1</td>
<td>GGCTGGGACCATGCGCGCCCTTCCCTTTTTTTGTTGTTACAGGCGATCTTGCA</td>
</tr>
<tr>
<td>MetKshort6A/RZ 3-1</td>
<td>GGCTGGGACCATGCGCGCCCTTCCCTTTTTTGTTGTTACAGGCGATCTTGCA</td>
</tr>
<tr>
<td>MetKshort88-89G/RZ 3-1</td>
<td>GGCTGGGACCATGCGCGCCCTTCCCTTTTTTGTTGTTACAGGCGATCTTGCA</td>
</tr>
<tr>
<td>MetKshortSmk/RZ 3-1</td>
<td>GGCTGGGACCATGCGCGCCCTTCCCTTTTTTGTTGTTACAGGCGATCTTGCA</td>
</tr>
<tr>
<td>MetKshortcr/RZ 3-2</td>
<td>GCCATCGATCTTCAATACGCTACATAGGGGTGATACACGCAACAAAG</td>
</tr>
<tr>
<td>MetKshortcr/RZ 3-3</td>
<td>GGCTGGGACCATGCGCGCCCTTCCCTTTTTTGTTGTTACAGGCGATCTTGCA</td>
</tr>
<tr>
<td>MetK-RNH4</td>
<td>ACAAGGCATC</td>
</tr>
</tbody>
</table>
3.2.2  *In vitro* SAM binding assays

We carried out SAM binding assays as previously described (Fuchs et al., 2006) with minor modifications. Full-length RNAs (3 μM) corresponding to positions 15-118 of the *metK* leader sequence were heated to 65 °C for 5 min in 1X transcription buffer (20 mM Tris-Cl, pH 8.0, 20 mM NaCl, 10 mM MgCl₂, 0.1 mM EDTA) followed by slow-cooling to 40 °C. Radiolabeled SAM (3 μM [methyl-³H]-SAM, 15 Ci mmol⁻¹; GE Healthcare) was added in a final reaction volume of 40 μl. Binding reactions were incubated at room temperature (22 °C) for 15 min followed by passage through a Nanosep Centrifugal Device (Pall Corporation) with a molecular weight cutoff (MWCO) of 10 kDa by centrifugation at 14,000 X g for 2.5 min. Filters were washed four times with 40 μl 1X transcription buffer to remove unbound SAM, and material retained by the filter was collected, mixed with Packard BioScience Ultima Gold scintillation fluid and counted in a Packard Tri-Carb 2100TR liquid scintillation counter. No SAM was retained by the filter in the absence of RNA; hence, nonspecific binding is negligible.

Truncated RNAs used for crystallization were tested for SAM binding using a similar protocol with the following modifications. Unbound SAM was removed from binding reactions by passage through a Nanosep Centrifugal Device (Pall Corporation) with a MWCO of 3 kDa by centrifugation at 14,000 X g. Wash steps were eliminated from this protocol to avoid possible dissociation of the RNA-SAM complex. The concentration of the RNA-[³H]-SAM complex retained by the filter and the unbound SAM present in the flow through was determined by scintillation counting of known volumes, and percent binding for truncated constructs was normalized to that observed.
with the full-length construct for comparison. All binding assays were carried out at least in duplicate.

### 3.2.3 RNase H cleavage assays

DNA templates were generated as described above and used as templates for RNase H cleavage experiments. Radiolabeled RNAs corresponding to *E. faecalis metK* leader RNA were synthesized using a MEGA-script T7 RNAP transcription kit (Ambion) in the presence of \([\alpha^{32}\text{P}]-\text{UTP}\). RNAs transcribed in the presence or absence of SAM (250 µM) were incubated with DNA oligonucleotide MetK-RNH4 (Table 3.1; 4.5 µM), which is complementary to positions 63-72, for 5 min at 37 °C followed by addition of RNase H (0.45 U ml\(^{-1}\); Ambion). Digestion was terminated after 10 min at 37 °C by addition of phenol and cleavage products were resolved on a 6% denaturing polyacrylamide gel. Digestion products were visualized by PhosphorImager analysis (Molecular Dynamics) and quantified using ImageQuant 5.2 software by determining the percentage of cleavage product relative to the total amount of both cleaved and full-length RNA. RNase H cleavage experiments were carried out in duplicate.

### 3.2.4 Apparent \(K_D\) determination

The apparent equilibrium dissociation constants (\(K_D\) values) for the wild-type 15–118 metK leader sequence and a 51-nucleotide truncated construct (SMK51) were determined using a modified SAM binding assay. Briefly, T7 RNAP transcribed RNA (1 µM) in 1X transcription buffer was refolded as described above and incubated with \[^3\text{H}\]-SAM ranging in concentration from 0.02 µM to 10 µM. Binding reactions were loaded
onto Nanosep Omega filters (Pall Life Sciences) with a MWCO of 3 kDa and centrifuged briefly at 14,000 X g to allow a small portion of the reaction to flow through. The concentration of the RNA-[³H]-SAM complex retained by the filter and the unbound SAM present in the flow through were determined by scintillation counting of known volumes, and nonlinear regression analyses were performed using KaleidaGraph Version 3.51 (Synergy Software). Apparent $K_D$ values represent the averages of at least two independent experiments for each construct with a margin of error $\leq 5\%$.

### 3.3 Results

#### 3.3.1 Design and testing of minimal $S_{MK}$ box constructs

Preliminary attempts to obtain diffracting crystals of the full-length *E. faecalis* $S_{MK}$ box riboswitch (corresponding to residues 15-118) were unsuccessful. Based on results obtained from previous mutational analyses (see section 3.1) we designed a series of truncated constructs in order to identify a minimal $S_{MK}$ box element that retains the ability to bind SAM (Fig. 3.2A). Each of the minimal constructs tested were truncated at both the 5’ and 3’ ends (initiating at residue 20 or 21 and ending at residue 94 or 95) and contained a five-bp stem capped with a GAAA-tetraloop in place of the hypervariable terminal stem loop (corresponding to residues 33-62). The hypervariable region upstream of the SD sequence (corresponding to residues 77-89) was replaced with a poly(A) linker and was systematically shortened to define the minimal sequence required for SAM binding.
Fig. 3.2. Design of a minimal SMK box element. A) Each of the minimal constructs considered for crystallography was truncated at both the 5’ and 3’ ends and contained a G/C rich stem and GAAA tetraloop (underlined) in place of the hypervariable stem at the top of the riboswitch. The ASD sequence and the complementary residues overlapping the SD sequence were maintained in each transcript (shaded). A series of constructs containing substitutions at residues 77-89 (boxed) were tested for their ability to bind SAM. B) SAM binding to minimal SMK box constructs was normalized to binding observed with wild-type RNA (15-118). Nucleotide sequences corresponding to residues 77-89 are indicated. SMK6 (used for crystallography) was constructed by the laboratory of Dr. Ke (Cornell University, Ithaca, NY). All other constructs and all binding assays were done by A.M. Smith. Data represent the averages from at least two independent experiments ± standard deviation. SMK51 and SMK6 are identical, except that SMK6 contains an additional G and U residue at the 5’ and 3’ ends respectively (not shown).

A size exclusion filtration assay was used to measure SAM binding by each of the minimal constructs (Fig. 3.2B). In several of the constructs (EfMetKshort8A, EfMetKshort7A, and EfMetKshort6A), the first two guanine residues of the SD sequence
(corresponding to residues 88 and 89) were substituted with adenine. These constructs exhibited a complete loss of SAM binding. Restoration of the nucleotides at positions 88 and 89 to guanine (in constructs EfMetKshort2G, SMK51, and SMK6) was sufficient to restore SAM binding to levels that were slightly higher than that observed with wild-type. These results indicate that the first two residues of the SD sequence are important for SAM binding, despite the fact that they do not participate in SD-ASD pairing.

3.3.2 RNase H cleavage analysis of *E. faecalis* met*K* transcripts

Accessibility of the core region of the SMK box riboswitch to an antisense DNA oligonucleotide was monitored by cleavage with RNase H (Fig. 3.3). Binding of oligonucleotide MetK-RNH4 complementary to residues 63-72 of the met*K* leader to the wild-type transcript (corresponding to residues 15-118) in the absence of SAM resulted in efficient cleavage by RNase H, consistent with previous results (Fuchs et al., 2006). Addition of SAM resulted in a ~3-fold decrease in the efficiency of RNase H cleavage, indicating that residues 63-72 of the met*K* leader are less accessible for interaction with the MetK-RNH4 oligonucleotide in the presence of SAM. These data support the hypothesis that residues 67-69 base pair with residues 29-31 when ligand is bound. A mutant construct (corresponding to residues 15-118) containing G88A and G89A substitutions also exhibited efficient oligonucleotide-directed RNase H cleavage in the absence of SAM (Fig. 3.3). However, no resistance to RNase H cleavage was observed upon SAM addition, indicating that residues 63-72 in the mutant construct do not become less accessible for interaction with the MetK-RNH4 oligonucleotide as in the wild-type met*K* leader. These results suggest that the G88A and G89A substitutions disrupt SAM-
dependent pairing interactions in the core element of the SMK box riboswitch, consistent with the observation that alteration of these residues disrupts SAM binding.

**Fig. 3.3. RNase H cleavage analysis of full-length and minimal SMK box transcripts.**

**A)** Denaturing PAGE analysis of RNase H cleavage products. Oligonucleotide MetK-RNH4, complementary to residues 63-72 (in the core element) was added to radiolabeled *E. faecalis* metK RNA generated in the presence (+) or absence (−) of SAM (250 µM). RNA-DNA hybrids were cleaved with RNase H and the products were visualized by autoradiography. Arrows indicate full-length transcripts (FL) and cleavage products (C) for 15-118 (left) and 51-nt constructs (right). **B)** Percentage of RNase H cleavage for each construct in the presence (gray bars) and absence (white bars) of SAM.

Accessibility of residues 63-72 in the core of the 51-nt minimal constructs EfMetKshort8A and EfMetKshort2G was also monitored using the RNase H cleavage assay (Fig. 3.3). Binding of the MetK-RNH4 oligonucleotide resulted in efficient RNase H cleavage of both of these minimal constructs in the absence of SAM. However, resistance to RNase H cleavage in the presence of SAM was observed only for the EfMetKshort2G construct. This observation is consistent with results from the SAM
binding experiments, which demonstrated that the EfMetKshort2G construct retains the ability to bind SAM while the EfMetKshort8A construct does not (Fig. 3.2). These results further demonstrate the importance of the first two residues of the SD sequence (G88 and G89) for SAM binding.

### 3.3.3 SAM binding affinity of the $S_{MK}$ box riboswitch and a 51-nt minimal construct

To verify the suitability of an $S_{MK}$ box minimal construct for crystallization studies, we performed SAM binding assays using increasing concentrations of SAM to determine an apparent binding affinity for the 51-nt construct SMK51. We also determined an apparent $K_D$ for the full-length construct for comparison. The $K_D$ values for SAM determined by size exclusion filtration were estimated to be 0.85 µM for the full-length transcript and 0.57 µM for SMK51 (Fig. 3.4). The $K_D$ value for the full-length transcript was further confirmed by fluorescence-quenching assays (see section 5.3.2). These results demonstrate that the 51-nt $S_{MK}$ box element exhibits SAM binding activity similar to that of the full-length construct. In addition, competition binding experiments revealed that the full-length $S_{MK}$ box riboswitch exhibits at least a 100-fold preference for SAM over SAH (see section 5.3.4).
The slightly higher binding affinity exhibited by the SMK51 RNA is likely to be due to the fact that this construct lacks residues at the 5’ end (corresponding to positions 15-19) that are predicted to stabilize the unliganded conformation. Therefore this RNA is predicted to favor the ligand-bound conformation even in the absence of SAM. Results from NMR spectroscopy experiments validate this hypothesis (see Chapter 4). Since the 51-nt construct exhibits a binding affinity that is at least comparable to that of the full-length RNA, we concluded that this minimal construct is suitable for crystallization experiments to determine the structure of the $S_{mk}$ box riboswitch in complex with SAM.
3.3.4 Crystal structure of the SMK box riboswitch

Crystal structure determination was performed by the laboratory of Dr. Ailong Ke (Cornell University, Ithaca, New York). The 53-nt minimal construct SMK6 was chosen for crystallographic analysis. This construct differs from SMK51 only by an additional residue at both the 5’ and 3’ ends. The 2.2-Å crystal structure of the SMK box riboswitch in complex with SAM revealed that this RNA folds into an inverted Y-shaped molecule (Fig. 3.5A and B). Helix P3 stacks on top of helix P2 to make up the long arm of the molecule, and helices P1 and P4 constitute the short arms. Most of the predicted secondary structural features of the SMK box riboswitch were verified by the crystal structure, including base pairing between residues in the SD sequence and the ASD sequence. Pairing interactions in the core of the molecule between residues 29-31 and 67-69 were also verified. The P4 helix, which involves pairing interactions between residues C76-C77 and residues G88-G89, was not previously assigned.
The crystal structure revealed that the ligand binding pocket of the $S_{MK}$ box riboswitch is located inside a three-way junction (at the intersection of helices P1, P2, and P4) and the adenosine moiety of SAM intercalates between U72 and G90 to allow continuous base stacking from helix P1 to P2. A base triple interaction (A73·G90-C25) defines the floor of the SAM binding pocket (Fig. 3.6A). This interaction tethers residue A73 (located in the P2/P4 junction) to the C25-G90 base pair that lies at the top of helix 90.
P1. A sheared base pair (U72·A64) along with a base triple (A27·G71·G66) form the ceiling of the SMK box ligand binding pocket (Fig. 3.6B). These tertiary interactions help to connect the J3/2 bulge to the base of helix P2. A conserved residue (U65) mediates a U-turn motif in the J3/2 bulge, which acts as a lid that encloses SAM in the electronegative environment of the ligand binding pocket.
Fig. 3.6. Key interactions in the SMK box ligand binding pocket. A) The base triple A73·G90-C25 forms the floor of the SAM binding pocket. B) Another base triple A27-G71-G66 defines the ceiling. C) Residues G26 and A73 form hydrogen bonds with the adenine moiety of SAM and D) residue G89 contacts the 2’ and 3’ hydroxyl groups of the SAM ribose. E) Residues G71 and U72 make favorable electrostatic interactions with the positive charge on the SAM sulfonium ion. The SAM adenosine moiety assumes a syn-conformation when bound to the SMK box riboswitch. F) The adenosine moiety of SAH assumes an anti-conformation in the ligand binding pocket with the sulfur and methionine moieties rotated 180º relative to the orientation of the same functional groups in SAM. Images were generated by the lab of Dr. Ailong Ke using coordinates entered into the RCSB Protein Data Bank under accession number 3E5C.
Several conserved residues in the $S_{MK}$ box ligand binding pocket make direct contacts with SAM. G26 forms three hydrogen bonds with the adenine moiety of SAM and A73 also makes a hydrogen bond with the adenine moiety (Fig. 3.6C). The $2'$ and $3'$hydroxyls of the ribose group of SAM make hydrogen bond interactions with residue G89 in the riboswitch RNA (Fig. 3.6D). Finally, the positively charged sulfonium ion of SAM makes favorable electrostatic interactions with the $2'$hydroxyl of residue G71 and the O4 carbonyl of U72 (Fig. 3.6E).

In contrast to the other two SAM binding riboswitches, no electron density was observed for SAM functional groups beyond the sulfonium ion indicating that the methionine tail is not specifically recognized by the $S_{MK}$ box riboswitch. The crystal structure also revealed that the methyl group attached to the sulfur ion is not specifically recognized by the $S_{MK}$ box element. Binding studies with SAM analogs validate these observations and are described in Chapter 5.

The adenosine moiety of SAM assumes an energetically unfavorable $syn$-conformation inside the $S_{MK}$ box ligand binding pocket, which orients its Watson-Crick face for hydrogen bonding to residue G26 of the riboswitch RNA (Fig. 3.6C). This conformation of the ligand is stabilized by intramolecular electrostatic interactions between the positively charged sulfonium group of SAM and the N3 and O4’ groups of the adenosine of SAM. The crystal structure of the $S_{MK}$ box riboswitch in complex with SAH (which lacks the positive charge on the sulfur) was investigated by displacement of SAM from the SMK6 crystals. These experiments revealed that the adenosine moiety of SAH assumes an $anti$-conformation when bound to the $S_{MK}$ box ligand binding domain (Fig. 3.6F). In addition, the favorable electrostatic interactions observed between the
sulfonium ion and $S_{MK}$ box residues G71 and U72 are lost when SAH is in place. These observations reveal that the positive charge on the SAM sulfur group is crucial for ensuring strong ligand-RNA interactions in the $S_{MK}$ box system. This is consistent with analog binding data that shows that the $S_{MK}$ box riboswitch binds SAM with 180-fold higher affinity than SAH (see section 5.3.3).

3.3.5 Mutational analysis of residues important for SAM binding

It was previously reported that base pairing between residues in the SD sequence and the ASD sequence is required for SAM binding (see section 3.1; Fuchs et al., 2006; Fuchs et al., 2007). However, it was unclear what other secondary structural features are required for SAM binding. We performed mutational analyses to investigate the importance of pairing interactions A29-U69, G30-C68, and G31-C67 in helix P2 of the $S_{MK}$ box riboswitch. Constructs containing single point mutations or double compensatory mutations were tested for their ability to bind SAM using a size exclusion filtration assay. All of the individual point mutations (A29G, G30A, G31A, C67U, C68U, and U69C), which are predicted to disrupt base pairing in the P2 stem, resulted in at least a 4-fold loss of SAM binding relative to wild-type RNA (Fig. 3.7). Restoration of base pairing in the G30A-C68U and A29G-U69C double mutants restored SAM binding to wild-type levels, while the compensatory mutation G31A-C67U only partially restored SAM binding. One possibility is that alteration of the G31-C67 pair is less tolerated for ligand binding because of the proximity of these residues to the J3/2 bulge, which contains a U-turn motif that acts as a lid enclosing SAM inside the ligand binding pocket (see section 3.3.4). This conclusion is supported by the observation that the G31-C67
residues are 100% conserved in all sequences analyzed, while the other pairing residues (G30-C68 and A29-U69) exhibit covariation, but are not 100% conserved (Fuchs et al., 2006). The results from these mutation experiments indicate that formation of these three base pairs in the P2 helix is essential for SAM binding.

Fig. 3.7. Mutational analysis of pairing residues in the P2 and P4 helices. Bars represent percent SAM binding (normalized to wild-type) to \( S_{MK} \) box constructs containing nucleotide substitutions as indicated. RNA transcripts were incubated with [methyl-\( ^3 \)H]-SAM, and unbound SAM was removed by size exclusion filtration. Data represent the averages of at least two independent experiments.

Binding assays with deletion mutations to identify a minimal \( S_{MK} \) box construct suitable for crystallography revealed the importance of residues G88 and G89 for SAM binding (see section 3.3.1). Crystallization of the SMK6 minimal construct showed that these residues pair with C75 and C76 to form helix P4 in the ligand-bound riboswitch. To verify the importance of these pairing interactions for ligand binding we performed
mutational analysis on residues C76 and G88. A construct containing a C76U substitution (that maintains a wobble base pair) retained 59% of SAM binding activity (relative to wild-type), while a G88A mutation that disrupts pairing resulted in a 75-fold reduction in SAM binding activity (Fig. 3.7). The compensatory double mutation C76U-G88A restored SAM binding to 32% of the wild-type level. These results suggest that pairing between residues 76 and 88 is important for SAM binding and that a weaker A-U pair at this position (in the context of the *E. faecalis metK* leader sequence) results in decreased stability of the P4 linker helix.

The crystal structure showed that residue G90 is involved in a base triple interaction (A73-G90-C25) that stabilizes the floor of the SAM binding pocket in the *S_mK* box riboswitch. As with other SD-ASD pairing interactions in the P1 helix, the C25-G90 base pair is important for SAM binding. A C25U mutant construct (which maintains a C25U-G90 wobble pair) retained 63% SAM binding activity while a G90A mutation that disrupts pairing exhibited a ~75-fold reduction in SAM binding activity relative to wild-type (Fig. 3.8). A C25U-G90A double mutation that restores base pairing was not sufficient to restore SAM binding activity. This is likely to be due to the loss of the N2 group when G90 is replaced with A, which disrupts the pairing with A73 in the base triple interaction. An A73G substitution in the riboswitch sequence, which also disrupts the A73-G90-C25 base triple, resulted in an 80-fold reduction in SAM binding activity. The requirement for the A73-G90-C25 base triple explains the absolute conservation of the SD sequence GGGGG (instead of the more commonly found GGAGG sequence) in all known *S_mK* box elements.
We also performed mutational analysis to investigate the importance of residue U65. The crystal structure revealed that this residue mediates a U-turn motif in the J3/2 bulge of the SMK box riboswitch that acts as a lid excluding SAM inside the ligand binding pocket. Substitution of the U65 by a C resulted in a 5-fold reduction in SAM binding activity, and deletion of this residue resulted in a 20-fold decrease in SAM binding activity (Fig. 3.8). These results demonstrate the importance of the U-turn motif for SAM binding to the SMK box RNA.

**Fig. 3.8. Mutational analysis of SAM binding site residues.** Bars represent percent SAM binding (normalized to wild-type) to SMK box constructs containing nucleotide substitutions as indicated. RNA transcripts were incubated with [methyl-3H]-SAM, and unbound SAM was removed by size exclusion filtration. Data represent the averages of at least two independent experiments. Binding assays for the C25U, G90A, and C25U G90A mutants were carried out by A.M. Smith and R.T. Fuchs. All other assays were done by A.M. Smith.
The crystal structure of SMK6 revealed that several RNA residues make direct contacts with SAM. In particular, G26 makes 3 hydrogen bond interactions with the adenine moiety of SAM. Substitution of G26 by an A resulted in a 5-fold decrease in SAM binding activity (Fig. 3.8). Also, the crystal structure revealed electrostatic interactions between U72 and the positive charge on the sulfonium ion of SAM. Substitution of U72 with a C, which places a partially positive N4 amine toward the SAM sulfonium group, resulted in a 70-fold decrease in SAM binding activity (Fig. 3.8). These results suggest that the favorable electrostatic interactions between U72 and SAM are crucial for ligand recognition. The S box and SAM-II riboswitches utilize a ligand recognition mechanism that also involves favorable electrostatic interactions between the positive charge on the sulfonium group of SAM and the O4 carbonyl of a uracil residue in the riboswitch sequence (Gilbert et al., 2008; Montange and Batey, 2006).

3.4 Discussion

We have defined a minimal SMK box element based on the *E. faecalis metK* leader sequence that retains the ability to bind SAM. The 53-nt RNA contains truncations at both the 5’ and 3’ ends, and deletions of residues in the hypervariable regions located at the top of the terminal stem loop and within the linker region upstream of the SD sequence. Despite these differences, the RNA retains the ability to bind SAM with an affinity that is similar to that of the full-length RNA. This construct was used to determine the 2.2-Å crystal structure of the SMK box riboswitch in complex with SAM. These data along with mutational analyses provide new insight into structural determinants required for binding of SAM to the riboswitch RNA. Most of the predicted
secondary structural features were verified by the crystal structure, and a new helix (P4) that was not previously assigned was identified. Based on these observations, a new secondary structure model was developed for the full-length *E. faecalis metK* leader RNA (Fig. 3.9).

**Fig. 3.9. Revised secondary structure of the *E. faecalis* SMK box riboswitch.** The crystal structure confirmed many of the predicted secondary structural features (see Fig. 3.1) of the SMK box riboswitch in complex with SAM (right). The only difference between the old and revised models is the assignment of residues 75–89. These residues form a helix that constitutes one arm (helix P4) of the riboswitch three-way junction.
The determination of the crystal structures of each of the SAM binding riboswitches has provided a comprehensive picture for the comparison of three RNAs that have evolved to recognize the same cellular metabolite. Each of these riboswitch elements adopts an entirely unique RNA fold for formation of the ligand binding pocket. In the S box riboswitch, SAM is encapsulated in a pocket between two helical stacks (Montange and Batey, 2006). The SAM-II riboswitch folds into a classic H-type pseudoknot with SAM located at the center of an RNA triple helix (Gilbert et al., 2008). Lastly, the SMK box riboswitch encloses SAM in a ligand binding pocket located at the center of a three-way junction (Lu et al., 2008).

The three SAM binding riboswitches also exhibit different ligand recognition properties. In the S box system, every functional group of the ligand besides the methyl is recognized through either direct or indirect contacts (Montange and Batey, 2006). In the SAM-II riboswitch, all of the SAM functional groups are important for ligand recognition (Gilbert et al., 2008). The crystal structure of the SMK box riboswitch revealed that the methionine tail and the methyl group attached to the SAM sulfur are not specifically recognized by the RNA. In all three cases, the adenine moiety of SAM is involved in base stacking and base triple interactions that are important for maintaining the ligand-bound conformation of the riboswitch RNA. Basic differences in the overall structures and recognition requirements of each class suggest that the three SAM binding riboswitches evolved independently during evolution.

While each of the SAM binding riboswitches exhibits a different mechanism for ligand recognition, they all utilize a similar strategy for the discrimination of SAM from the naturally occurring analog SAH. Each of the three systems utilizes favorable
electrostatic interactions between the positively charged sulfonium ion of SAM and O4 carbonyls of conserved uridine residues in the RNA sequence for ligand recognition. Each of the classes also exploits favorable intramolecular electrostatic interactions within the SAM molecule itself to stabilize a syn-conformation of the ligand adenosine moiety. SAH, which lacks the positive charge on the sulfur, does not exhibit these intramolecular interactions and the ligand adenosine moiety instead favors an anti-conformation. Because each of these riboswitches recognizes functional groups in both the adenosine and sulfonium moieties simultaneously, they reject SAM analogs (such as SAH) that are unable to assume the syn-conformation. The loss of the positive charge on the sulfur atom appears to be the primary feature responsible for the reduction in SAH binding affinity (>100-fold lower than for SAM) observed with each of these riboswitch classes. While the three SAM binding riboswitches appear to have evolved independently, they have converged at the functional level to specifically recognize SAM and reject SAH by exploiting the modest structural differences between the two analogs.

The SAH binding riboswitch is another RNA that is able to distinguish between the two closely related compounds (Wang et al., 2008). This riboswitch selectively binds SAH with ~1000-fold higher affinity than SAM. Analog studies with the SAH riboswitch revealed that RNA interaction with virtually every ligand functional group is required for optimal recognition. Additional structural studies are needed to characterize specific interactions and to determine the conformation of SAH inside the ligand binding pocket for comparison to the SAM binding riboswitches.

Most riboswitches are modular in architecture, consisting of a ligand binding aptamer domain and a separate regulatory output domain (e.g., the THI box and S box
riboswitches). Typically, ligand binding at the aptamer domain triggers a structural rearrangement of the output domain that affects the efficiency of expression of the downstream ORF. The SMK box riboswitch exhibits a much simpler architecture in which the regulatory domain is part of the ligand binding domain (Fig. 3.10). This means that ligand binding and the regulatory event (i.e., sequestration of the SD sequence) occur in a single step. Only one other riboswitch class, the SAM-II riboswitch, is predicted to utilize a similar architecture comprised of a single domain. However, the mechanism of regulation utilized by the SAM-II riboswitch has not been verified. Therefore, the SMK box riboswitch is the first riboswitch class identified in which residues that are required for ligand recognition (the SD sequence) are also directly involved in gene regulation.
Riboswitches that have a modular architecture can couple the ligand binding
domain to a variety of regulatory output domains. For instance, members of the THI box
riboswitch class, which exhibit a conserved domain for ligand binding, have evolved
several distinct mechanisms of regulation (see section 1.2.10). The $S_{MK}$ box riboswitch,
on the other hand, may be less tolerant to sequence variations, since the same residues
that are required for regulation are also essential for ligand binding. Alternatively, it may
be possible for the $S_{MK}$ box riboswitch to function in the context of a modular riboswitch.
architecture where sequence motifs in helices P1 and P4 still participate in ligand binding but no longer constitute the regulatory platform.
CHAPTER 4

INVESTIGATION OF ALTERNATE CONFORMATIONS OF THE SMK BOX RIBOSWITCH USING NMR SPECTROSCOPY

4.1 Introduction

All metabolite binding riboswitches characterized to date utilize a mechanism of regulation that involves structural rearrangement of the RNA in response to a molecular signal. Each riboswitch RNA is predicted to favor one of two alternate conformations depending on the presence or absence of its cognate molecular signal. The conformational state of the riboswitch determines the efficiency of expression of a cis-encoded ORF. For example, in the absence of ligand, the SMK box riboswitch is predicted to favor a conformation in which the SD sequence is accessible allowing for efficient translation of \textit{metK}. Binding of SAM to the riboswitch leads to stabilization of an alternate conformation that exhibits decreased accessibility of the SD sequence, resulting in decreased translation initiation.

The crystal structure of a 53-nt minimal SMK box element in complex with SAM revealed structural features of the ligand-bound conformation of this riboswitch (see Chapter 3; Lu et al., 2008). These studies showed that the RNA forms a Y-shaped molecule with SAM located in the center of a three-way junction at the intersection of
helices P1, P2, and P4. Pairing interactions in each of these helical domains is required for SAM binding, and formation of helices P1 and P4 results in sequestration of residues 88-92 that constitute the \textit{metK} SD sequence, providing a structural basis for SAM-dependent translational repression. These results verified the prediction that the SMK box riboswitch assumes a simplistic architecture comprised of a single domain that is responsible for both ligand binding and regulatory control (see section 3.4). This is in contrast to most other riboswitch RNAs, which are comprised of two separate domains.

Less evidence is available about the structure of the unliganded conformation of the SMK box riboswitch. It is predicted based on phylogenetic analysis that in the unliganded conformation helices P1, P2, and P4 are absent and that another helix (referred to here as P0) is formed. Helix P0 is predicted to be comprised of residues 13-20 (located upstream of the ASD sequence) and residues 68-75, which include nucleotides that overlap portions of helices P2 and P4 in the ligand-bound conformation. Therefore, the two alternate conformations (ligand-bound and unbound) are mutually exclusive and stabilization of helices P2 and P4 upon SAM binding results in disruption of helix P0. Results from a variety of experiments (e.g., enzymatic probing, RNase H cleavage assays, 30S ribosomal toeprinting) provide strong evidence that the RNA assumes an alternate conformation in the absence of ligand and exhibits secondary structural features consistent with the model (Fuchs et al., 2006, 2007).

In the current study we have used NMR spectroscopy to investigate pairing interactions of the SMK box riboswitch in the presence and absence of SAM. NMR experiments were conducted in collaboration with the laboratory of Dr. Mark Foster (The Ohio State University, Columbus, OH). We show that a 51-nt minimal construct
(SMK51) lacking residues 13-20 favors the ligand-bound conformation (consistent with the crystal structure) even in the absence of SAM. Addition of residues 13-20 to generate a 59-nt construct (SMK59) was sufficient to restore the ability of the riboswitch RNA to assume two alternate conformations depending on the presence of SAM. These results validate the predicted secondary structural features of the SMK box riboswitch in each conformational state and demonstrate that residues at the 5’end of the element (which are not required for ligand binding) are essential for stabilizing the ligand-free conformation. These data represent the first atomic-level structural evidence of an unliganded riboswitch using an RNA element that encompasses both the ligand binding domain and regulatory platform.

4.2 Materials and Methods

4.2.1 RNA synthesis and purification

The DNA template for T7 RNAP transcription of SMK51 was constructed by performing sequential PCR reactions using oligonucleotides (Integrated DNA Technologies, Coralville, IA) containing partially overlapping sequences for each successive round. SMK59 was constructed similarly except that SMK51 DNA was used as a starting template and the modified 5’end was incorporated using sequential PCR reactions. DNA oligonucleotides used for synthesis of SMK51 and SMK59 are shown in Table 3.1 and Table 4.1, respectively. Both of the minimal constructs used for NMR spectroscopy contained deletions in the SMK box riboswitch hypervariable regions as described in Chapter 3. Constructs were flanked by a hepatitis delta virus ribozyme at the 3’ end and a hammerhead ribozyme at the 5’ end to ensure homogenous termini.
(corresponding to the desired sequence) and were inserted into a pUC18 plasmid using EcoRI and BamHI restriction endonucleases (New England Biolabs). A T7 RNAP promoter sequence was included upstream of the hepatitis delta ribozyme to facilitate transcription initiation. Additional sequence was inserted between the T7 RNAP promoter and the ribozyme sequence to ensure that the upstream fragment produced by ribozyme cleavage could be easily separated from the SMK box riboswitch sequence using denaturing PAGE. DNA templates were linearized by digestion with the restriction endonuclease HindIII (New England Biolabs).

Transcription reactions (30 ml) contained 40 mM Tris-Cl, pH 8.0, 22 mM MgCl₂, 2 mM spermidine, 5 mM DTT, 50 µg ml⁻¹ BSA, 4 mM GTP, 4 mM ATP, 4 mM UTP, 4 mM CTP, 20 mM GMP, 700 µg linearized template DNA, 8 units ml⁻¹ RNase inhibitor (Roche), 2 µg ml⁻¹ pyrophosphatase (Roche), and 45 µg ml⁻¹ T7 RNAP and were incubated overnight at 37 °C. For ribozyme cleavage, samples were diluted 5-fold, the concentrations of MgCl₂ and Tris-Cl, pH 8.0 were increased to 20 mM and 50 mM, respectively, and samples were incubated at 50 °C for 1 hour (based on verbal communication from Herve Roy, The Ohio State University, Columbus, OH). RNA was ethanol precipitated and dried pellets were stored at -20 °C until PAGE purification.
Transcription of isotopically-labeled RNAs was carried out identically except that NTPs were replaced with $^{15}$N- and $^{13}$C-labeled NTPs (Isotech, Sigma Aldrich).

RNAs corresponding to S_{MK} box riboswitch constructs were separated from ribozyme fragments using 10% denaturing PAGE. RNA bands were excised and electroeluted in 1X TBE followed by phenol:chloroform extraction and ethanol precipitation. RNAs were resuspended in 5 ml of chromatography Buffer A (20 mM Tris-Cl, pH 7.5, 0.5 mM EDTA) and loaded onto a HiTrap DEAE FF anion exchange chromatography column (GE Healthcare) equilibrated with Buffer A. The flow-through was collected and reloaded onto the chromatography column two additional times. The column was rinsed with 25 ml of Buffer A and RNA retention by the resin was verified by measuring the A$_{260}$ of the flow-through and wash fractions. Elution was carried out over 25 min (2 ml min$^{-1}$ flow rate) using a stepwise gradient in which Buffer A and Buffer B (20 mM Tris-Cl, pH 7.5, 0.5 mM EDTA, 1M NaCl) were gradually mixed using a Pharmacia LKB fast protein liquid chromatography system (FPLC). Fractions (1 ml) were collected and RNA content was monitored by measuring the A$_{260}$ and by denaturing PAGE. RNA fractions were combined and ethanol precipitated, and dried pellets were resuspended in a minimal volume (0.5-1.0 ml) of 25 mM potassium phosphate buffer, pH 6.2, 50 mM NaCl. RNAs were dialyzed three times against 1X NMR buffer (25 mM potassium phosphate buffer, pH 6.2, 50 mM NaCl, 5 mM MgCl$_2$, 10% D$_2$O) and stored at -20 ºC.

All NMR experiments were performed by Ross Wilson in the laboratory of Dr. Mark Foster (The Ohio State University, Columbus, OH). Prior to data collection, SMK51 and SMK59 RNA samples were annealed by heating to 65 ºC for five minutes.
and slow cooling to 4, 15, or 25 °C to allow for proper folding. RNA samples were quantitated by measuring the $A_{260}$ and were typically estimated to be ~500 µM. The NMR spectra were recorded at 4, 15, or 25°C using 600 and 800 MHz Brüker Avance DRX spectrometers (Billerica, MA) equipped with triple resonance pulse-field gradient probes. NMR spectra were processed and analyzed using NMRPipe.

4.2.2 In vitro SAM binding assays

SAM binding assays were performed as described in section 3.2.2. Briefly, RNAs (1 µM) in 1X transcription buffer (20 mM Tris-Cl, pH 8.0, 20 mM NaCl, 10 mM MgCl₂, 0.1 mM EDTA) were refolded by heating to 65 °C for 5 min followed by slow cooling to 40 °C. Radiolabeled SAM (1 µM [methyl-³H]-SAM, 15 Ci mmol⁻¹; GE Healthcare) was added in a final reaction volume of 40 µl and binding reactions were incubated at room temperature (22 °C) for 15 min. Unbound SAM was removed from binding reactions by passage through a Nanosep Centrifugal Device (Pall Corporation) with a MWCO of 3 kDa by centrifugation at 14,000 X g. The concentration of the RNA-[³H]-SAM complex retained by the filter and the unbound SAM present in the flow through was determined by scintillation counting of known volumes, and percent binding for truncated constructs was normalized to that observed with the full-length construct. Binding assays were carried out at least in duplicate.
4.3 Results

4.3.1 Synthesis and purification of SMK51 RNA

Initial NMR spectroscopy experiments were performed using a minimal SM box riboswitch element that was originally developed as a candidate for crystallography experiments (Fig. 4.1). This construct, called SMK51, initiates at the first residue of the ASD sequence (U21) and terminates at the last residue of the SD sequence (A94). SMK51 was shown using a size exclusion filtration assay to bind SAM with slightly higher affinity than the full-length construct corresponding to residues 15-118 of the E. faecalis metK transcript (apparent $K_D$ values were 0.57 and 0.85 µM respectively; see Fig. 3.4).
Fig. 4.1. Predicted secondary structure of SMK51 and SMK59 used for NMR spectroscopy. The 51 and 59 nucleotide constructs are identical except that SMK59 has eight additional residues at the 5’end (shown in gray) that are missing from SMK51. These residues are predicted to be important for stabilizing the unliganded conformation (left) of the SMK box riboswitch. Key pairing regions that stabilize helices P1, P2, and P4 in the presence of SAM (right) are color coded as in previous Chapters.

A hepatitis delta virus ribozyme was inserted downstream of the SMK box element to avoid 3’ end heterogeneity. Additionally, a hammerhead ribozyme was included at the 5’end so that the SMK box riboswitch element could be initiated with a U residue (and not with a G residue as required for efficient T7 RNAP transcription initiation). A T7 RNAP promoter element, followed by several G residues, was inserted upstream of the 5’hammerhead ribozyme. Ribozyme self-cleavage occurred spontaneously during transcription and RNA products loaded onto a 10% denaturing polyacrylamide gel migrated as several distinct bands that are likely to correspond to the
full-length transcript, cleaved ribozyme fragments, partially cleaved products, and the
fully processed $S_{MK}$ box construct (Fig. 4.2A).

Fig. 4.2. Large-scale purification of $S_{MK}$ box riboswitch for NMR spectroscopy. A) SMK51 (pictured) and SMK59 fragments were separated from ribozyme cleavage products using denaturing PAGE (10% acrylamide, 8M urea). Residual acrylamide was removed by passage over an anion exchange column. RNA was eluted using a NaCl gradient and fractions were analyzed by B) measurement of the A260 and C) by denaturing PAGE.

Products from large-scale transcription reactions (30 ml) were purified using
denaturing PAGE. The fragment corresponding to the fully processed $S_{MK}$ box riboswitch element was excised from the gel and electroeluted. Preliminary NMR spectroscopy
experiments revealed that contaminating acrylamide was present in the sample even after phenol:chloroform extraction (not shown). Therefore, we used anion exchange chromatography to remove residual acrylamide from subsequent RNA preparations. RNA was eluted from the column using an NaCl gradient and fractions were analyzed using absorbance spectroscopy (Fig. 4.2B) and denaturing page (Fig. 4.2C). The eluted RNA fragment migrated as a doublet; this is likely to be due to incomplete denaturation of the sample. RNA fractions were pooled and concentrated using ethanol precipitation and resuspended in a potassium phosphate buffer, pH 6.2 for NMR analysis. Attempts to purify RNA directly using anion exchange were unsuccessful as SMK box riboswitch fragments could not be resolved from other ribozyme cleavage products. Therefore, it was necessary to perform the PAGE purification step prior to anion exchange chromatography for each subsequent RNA preparation.

4.3.2 1D imino proton spectra of SMK51

All NMR spectroscopy experiments were carried out by Ross Wilson in the laboratory of Dr. Mark Foster (The Ohio Sate University, Columbus, OH). 1D imino proton spectra of the SMK51 construct were initially obtained with RNA suspended in 25 mM potassium phosphate buffer (pH 6.2) containing 50 mM NaCl in the absence of Mg$^{2+}$ (Fig. 4.3A). The number of detectable signals in the imino proton range corresponding to Watson-Crick base pairs (between 12.0-14.0 ppm) decreased from ~12 to ~4 as the temperature was increased from 15 °C to 36 °C, presumably because of melting of the RNA structure. Some peaks observed in the absence of Mg$^{2+}$ (particularly at lower
temperatures) exhibited broad linewidths, which can be induced by exchange between multiple conformations present in the sample.

![Fig. 4.3. 1D imino proton spectra of SMK51. A) RNA was resuspended in 25 mM phosphate buffer, pH 6.2, 50 mM NaCl and spectra were recorded at 15, 25, and 36 °C at 800 MHz. B) MgCl2 (5 mM) was added to samples suspended in buffer as in A) and spectra were recorded at 25 °C with increasing concentrations of SAM. Values to the right of the spectra represent the molar-equivalents of SAM added to each sample (relative to RNA). Figure provided by Ross Wilson and Mark Foster (The Ohio State University, Columbus, OH).](image)

Addition of MgCl2 resulted in more clearly defined spectra with imino peaks that exhibited narrower lineshape (compare Fig. 4.3A [25 °C] to Fig. 4.3B [none]). This suggests that magnesium ions result in increased homogeneity of the riboswitch conformational state, consistent with the hypothesis that divalent cations are required for tertiary folding of the RNA. 15 divalent metal ions were identified in the crystal structure of the SMK box riboswitch that are thought to be important for countering unfavorable
electrostatic interactions between sugar-phosphate backbones during RNA folding (Lu et al., 2008). SAM binding assays revealed that Mg$^{2+}$ is essential for ligand binding and that optimal binding occurs in reactions containing 2-10 mM MgCl$_2$ (data not shown). Based on these results, all subsequent NMR spectra were obtained with RNA resuspended and dialyzed in potassium phosphate buffer containing 5 mM MgCl$_2$.

1D imino proton spectra of SMK51 in the presence of SAM were also collected (Fig. 4.3B). Spectra obtained in the presence of SAM exhibited several new or enhanced imino peaks, but otherwise resembled spectra obtained in the absence of SAM. This effect was shown to be concentration-dependent as the intensity of several peaks increased upon addition of a higher concentration of SAM (compare 0.5- and 1.1-molar equivalents of SAM relative to the RNA concentration, Fig. 4.3B). These results suggest that some of the Watson-Crick base pairs become more structured in response to SAM binding, consistent with the hypothesis that SAM-mediated interactions help to stabilize folding of the ligand-bound structure.

4.3.3 2D NOESY spectra of SMK51 RNA in the presence and absence of SAM

We also obtained $^1$H-$^1$H NOESY (nuclear Overhauser effect spectroscopy) spectra of the SMK51 construct in the presence and absence of SAM (Fig. 4.4). These spectra are used to identify imino proton resonances from nuclear spins that are in close proximity to one another (<5 Å). The signals located along the diagonal correspond to individual imino proton signals (as in the 1D imino spectra described above), and the symmetrical crosspeaks that fall outside of the diagonal connect resonances from imino protons that are spatially close. Due to the 3.4 Å spacing between base pairs in a helix,
each helical imino is capable of transferring its magnetization to its immediate neighbor imino protons in either direction (in the same strand or the paired strand) but not to imino protons that are more distant. In the case of Watson-Crick base-pairing (where there is one imino from either U or G per base pair), a helix will result in a spectrum with a linear connection between imino signals along the helix. This makes NOESY connectivity of iminos an effective tool for identification of helical regions in RNA. An increase in the number of clearly defined imino proton signals (along the horizontal) was observed upon SAM addition (compare Fig. 4.4A and B), consistent with results from the 1D imino proton spectra described above (see section 4.3.2). In addition, there are fewer clearly defined crosspeak signals, indicative of base pairs that are spatially close to one another, in the absence of SAM. These results suggest that SAM binding results in stabilization of some base pairing interactions in SMK51 RNA, consistent with the model.
Fig. 4.4. 2D $^1$H-$^1$H NOESY spectra of SMK51 RNA in the presence and absence of SAM. Spectra were collected in 25 mM potassium phosphate buffer, pH 6.2, 50 mM NaCl, and 5 mM MgCl$_2$ at 4 ºC. Crosspeaks (connected by dashed lines) are used to identify imino proton signals that are in close proximity to one another. Sequential NOE contacts (depicted with colored arrows) revealed two helical regions in the absence of SAM (left) and four helical regions in the presence of SAM (right). Figure provided by Ross Wilson and Mark Foster (The Ohio State University, Columbus, OH).

By investigating the sequential imino-imino NOE contacts, helical regions in the SMK51 structure can be identified. In the absence of SAM, two distinct helical regions were observed (Fig. 4.4A), while in the presence of SAM, four distinct helical regions were observed (Fig. 4.4B). The four helical regions in the SAM-bound conformation are likely to correspond to helices P1 through P4, which were shown to be present in the crystal structure of the 53-nt construct SMK6 (see Chapter 3). The 5-imino connectivity (indicated by black arrows in Fig. 4.4) observed both in the presence and absence of SAM is likely to correspond to the P3 helix of the $S_{\text{MK}}$ box riboswitch. This is supported by the observation that the peak at 10.8 ppm (at one end of the 5-imino connectivity) has an extreme chemical shift (<11 ppm) characteristic of a G imino
participating in a sheared G·A mismatch as is found in a tetraloop (Rudisser and Tinoco, 2000). The identity of the other helical regions identified in the 2D NOESY spectra (two of which are observed only in the presence of SAM) is less evident. Although the SMK51 construct appears to lose some secondary structural features in the absence of SAM, there is no evidence for the formation of an alternate helix (i.e., helix P0) in the unliganded conformation. This is consistent with the hypothesis that residues located at the 5’end of the SMK box riboswitch, which are absent from SMK51, are required for formation of helix P0. This observation led to the design of a second construct (SMK59) containing additional residues at the 5’end in an attempt to verify the conformational state of the SMK box riboswitch in the absence of SAM.

4.3.4 Binding of SAM to the SMK59 transcript

SMK59 is identical to SMK51 except that it contains eight additional residues at the 5’end (corresponding to positions 13-20) that are predicted to be required for formation of helix P0 in the SMK box riboswitch (Fig. 4.1). SMK59 RNA was purified using denaturing PAGE and anion exchange chromatography as described for SMK51. To test the ability of SMK59 RNA to bind SAM, we used a size exclusion filtration assay (Fig. 4.5). We included full-length E. faecalis metK leader RNA (corresponding to residues 15-118) and SMK51 as positive controls. We also included a mutant construct containing a G88A substitution as a negative control. The G88A mutation was shown previously to result in a 75-fold reduction in SAM binding activity (see section 3.3.5). Consistent with previous results, the SMK51 construct bound SAM with 150% activity relative to the 15-118 construct. This increase in binding activity is likely to be due to the
absence of residues 15-20 that are predicted to stabilize the unliganded conformation of
the $S_{MK}$ box riboswitch. The absence of these residues is predicted to drive the RNA
equilibrium towards the ligand-bound conformation. This hypothesis is supported by the
observation that $S_{MK}59$ RNA, which includes residues 13-20, exhibits binding activity
similar to wild-type. Based on these results, $S_{MK}59$ is predicted to undergo
conformational switching in response to SAM binding that is similar to the full-length
riboswitch.

**Fig. 4.5. SAM binding to $S_{MK}59$ using size exclusion filtration.** Bars represent
percent SAM binding (normalized to wild-type) to $S_{MK}$ box constructs indicated on the x-
axis. RNA transcripts were incubated with [methyl-$^3$H]-SAM, and unbound SAM was
separated by size exclusion filtration. Data represent the averages from at least two
independent experiments ± standard deviation.

**4.3.5 2D NOESY spectra of $S_{MK}59$ RNA in the presence and absence of SAM**

The 2D $^1$H-$^1$H NOESY spectrum of the $S_{MK}59$ construct in the presence of
SAM exhibited a pattern that was identical to that observed with $S_{MK}51$ (Fig. 4.6B).
Four helical regions were identified based on sequential imino-imino NOE contacts, consistent with the formation of helices P1-P4 in the ligand-bound structure. In contrast, the NOESY spectrum of SMK59 RNA in the absence of SAM revealed a new pattern of imino proton signals (Fig. 4.6A). As with SMK51, the sequential NOE contacts presumed to correspond to the P3 stem at the top of the SMK box riboswitch remained intact (indicated by black arrows in Fig. 4.6). However, a new helical region was identified in the absence of SAM that is predicted to correspond to base paired residues in helix P0 of the unliganded RNA (indicated by pink arrows in Fig. 4.6A). These results verify that SMK59 RNA undergoes a conformational rearrangement upon ligand binding that involves disruption of one helix and formation of several other helical regions, consistent with the model. In both SMK51 and SMK59 RNAs, helix P3 at the top of the terminal stem loop appears to form regardless of the presence or absence of SAM.
Fig. 4.6. 2D $^1$H-$^1$H NOESY spectra of SMK59 RNA in the presence and absence of SAM. Spectra were collected in 25 mM potassium phosphate buffer, pH 6.2, 50 mM NaCl, and 5 mM MgCl$_2$ at 4 °C. Sequential NOE contacts (depicted with colored arrows) revealed two helical regions in the absence of SAM (left), including one (indicated with pink arrows) that was not observed with SMK51 RNA and that is predicted to correspond to helix P0 in the unliganded RNA. Four helical regions were observed with SMK59 RNA in the presence of SAM (right) that were identical to those observed with SMK51 RNA. Figure provided by Ross Wilson and Mark Foster (The Ohio State University, Columbus, OH).

4.3.6. Resonance assignment using HSQC NMR spectroscopy

To assign the imino proton resonances observed with the SMK59 construct in the presence and absence of SAM, we synthesized isotopically labeled RNA by incorporating $^{15}$N- and $^{13}$C-labeled NTPs during RNAP transcription. The imino proton signals were then assigned using $^{15}$N-$^1$H HSQC (heteronuclear single quantum coherence) NMR spectroscopy. This experiment reports the chemical shift of the nitrogens attached to the imino protons (previously observed in the $^1$H experiments), which exhibit distinctive $^{15}$N chemical shifts, allowing assignment as either G or U.
Specifically, a guanine N₁ has a chemical shift <155 ppm while a uracil N₃ appears at >155 ppm.

The five sequential NOE contacts observed with SMK59 RNA in the presence of SAM (indicated by black arrows in Fig. 4.6B) were assigned as five consecutive guanine residues, consistent with the hypothesis that these imino protons correspond to helix P3 at the top of the RNA (Fig. 4.7B). The NOE connectivities that define the 3-bp helical region (indicated by blue arrows in Fig. 4.6) were shown to correspond to consecutive GGU imino protons, and the two other helical regions (indicated by red and green arrows in Fig. 4.6) were shown to correspond to consecutive GG imino protons (Fig. 4.7B). Although these assignments are consistent with the formation of helices P1, P2, and P4 in the ligand-bound conformation, we cannot definitively assign these resonances to individual base pairs because of similarities between the secondary structures of each helix. For example, consecutive GG resonances could arise from the imino protons of any of the three helical domains (P1, P2, or P4), and the connected GGU resonances could correspond to either helix P1 or P2.
Fig. 4.7. Assignment of imino proton resonances of SMK59. Top panels show the $^1$H-$^1$H NOESY spectra of SMK59 in the presence (right) and absence (left) of SAM (as in Fig. 4.6). Bottom panels show the $^{15}$N-$^1$H HSQC spectra, which allow for the assignment of the corresponding imino proton signals. Imino resonances associated with guanine residues are shifted up relative to those associated with uracil residues. The high level of background noise in the SMK59 spectra in the presence of SAM is likely to be due to the 5’ residues that are predicted be unstructured in the ligand-bound conformation. Figure provided by Ross Wilson and Mark Foster (The Ohio State University, Columbus, OH).

The $^{15}$N-$^1$H HSQC spectra of SMK59 RNA in the absence of SAM confirmed that four sequential NOE contacts that comprise one of the two observed helical regions correspond to the guanine imino protons in helix P3 (Fig. 4.7A). The fifth sequential NOE resonance, which is present in spectra obtained in the presence of SAM, is missing from spectra obtained in the absence of ligand. This is likely to be due to destabilization of base pairs at the bottom of the P3 helix when the RNA assumes the unliganded conformation. The second helical region observed in the unliganded SMK59 RNA was
shown to correspond to base pairs in the P0 helix of the SMK box riboswitch with resonances produced by consecutive GUUU imino protons (Fig. 4.7A). There are two possible arrangements for the sequential GUUU imino protons in the P0 helix. The resonance of the G imino proton could correspond to either G13 or G71. Model helices of the P0 stem were used to test the directionality of the GUUU assignment (data not shown). 1H-1H NOESY spectra of the model helices verified that the observed imino protons correspond to residues G71, U72, U15, and U14 in the P0 helix. Additional base pairs at the terminal ends of helix P0 are probably not observable because of conformational exchange that prevents detection of the imino protons due to their increased exposure to solvent. Taken together, these results validate the secondary structural model of the SMK box riboswitch and show that in the absence of SAM helices P1, P2, and P4 are missing and that an alternate conformation is favored that involves formation of helix P0.

4.4 Discussion

Previous crystallographic analysis revealed the high-resolution structure of the SMK box riboswitch in complex with SAM (see Chapter 3; Lu et al., 2008). However, only limited evidence was available in support of the proposed unliganded conformation of the RNA. In the current study we used NMR spectroscopy to investigate the secondary structure of the SMK box riboswitch in the presence and absence of SAM. NMR spectroscopy was performed using two different minimal constructs. The first construct, SMK51, lacks residues at the 5’ end (as in the SMK6 crystallography construct) that are predicted to stabilize the unliganded conformation of the riboswitch RNA (Fuchs et al.,
Our results confirmed that this RNA folds into a structure that includes four helical regions in the presence of SAM, consistent with the crystal structure data. Although two fewer helices were observed in the absence of SAM, the data revealed that SMK51 RNA is unable to form an alternate competing structure (that involves formation of a new helix) when ligand is not present due to the absence of residues important for the formation of helix P0. Instead, it appears that the ligand binding pocket of SMK51 RNA is partially preformed and that SAM binding mediates tertiary interactions that stabilize the ligand-bound conformation of the RNA.

SMK59 RNA, which contains eight additional residues at the 5’end, was shown to undergo a more complete conformational switch in response to SAM binding, consistent with the model. In the absence of SAM, the RNA forms the P0 helix as well as the terminal stem-loop at the top of the riboswitch that constitutes the P3 helix. These results validate the proposed unliganded conformation of the SMK box riboswitch. Binding of SAM to SMK59 RNA results in disruption of the P0 helix and stabilization of the P1, P2, and P4 helices. These data show that the solution structure of the SMK box riboswitch in complex with SAM is consistent with the structure of the ligand-bound riboswitch as determined by x-ray crystallography.

Nearly all high-resolution structural studies involving riboswitches (e.g., the S box, G box, THI box, etc.) have used RNA sequences that include the ligand binding domain but not the regulatory element. While these studies provide structural insight into the residues required for ligand binding, they do not provide a complete picture of the regulatory mechanism of the riboswitch and the interplay between the ligand binding and regulatory domains. Also, these studies fail to provide insight into the unliganded
conformation of riboswitch RNAs because structural features that are predicted to be important for stabilizing the ligand-free structure are also missing. For instance, NMR studies with the G box riboswitch in the *B. subtilis* xpt-*pbuX* gene were performed using a transcript corresponding only to the ligand binding domain (Ottink et al., 2007). The RNA used in these studies is missing the downstream terminator element as well as residues that are required for formation of a competing antiterminator element (and subsequent disruption of helix P1 at the base of the ligand binding domain) in the absence of ligand. Results from these studies suggest that the ligand binding domain by itself is largely preformed to bind guanine. It is predicted that as the ligand binding platform is transcribed, it assumes a prefolded architecture that is stabilized by guanine, allowing for formation of a downstream terminator helix. Alternatively, if guanine is not present, as transcription proceeds past the ligand binding platform the RNA is predicted to undergo a structural rearrangement that involves disruption of helix P1 at the base of the ligand binding domain and formation of an antiterminator element. However, no high-resolution structural data are available to validate the conformational state of the G box riboswitch (either in the presence or absence of guanine) using an RNA element that extends beyond the ligand binding platform.

The compact single domain architecture of the SMK box riboswitch has allowed us to obtain high-resolution structural data using RNA constructs that encompass both the ligand binding domain and the regulatory platform. Our NMR spectroscopy data confirm that the riboswitch RNA assumes one of two mutually exclusive conformations depending on the presence or absence of SAM. These results provide the first high-resolution structural data for both conformational states of a riboswitch RNA and provide
a more complete picture of the regulatory mechanism utilized by the $S_{MK}$ box riboswitch. In contrast to the G box riboswitch described above, the $S_{MK}$ box riboswitch is not predicted to be preformed for ligand binding during transcription. As the mRNA is synthesized, elements required for formation of helix P0 become available before those that are required for formation of helices P1, P2, and P4. Unless SAM binding occurs, the $S_{MK}$ box riboswitch maintains the unliganded conformation in which the ligand binding domain is not intact and gene expression is turned on. This ensures that the $metK$ transcript is poised for translation initiation of the downstream coding region until SAM binding occurs, causing a conformational switch of the RNA. This translational mechanism of regulation of the $metK$ gene, in which the default state is “on,” allows for the rapid synthesis of SAM, an essential methyl donor that plays a central role in many different metabolic pathways.
CHAPTER 5

MEASUREMENT OF LIGAND BINDING TO THE SMK BOX RIBOSWITCH USING FLUORESCENCE SPECTROSCOPY

5.1 Introduction

The SAM binding riboswitches provide a unique opportunity to directly compare the recognition properties of three naturally occurring RNAs that have evolved to recognize the same ligand. Determination of the high resolution crystal structures of the S box, SAM-II, and SMK box riboswitches (Gilbert et al., 2008; Lu et al., 2008; Montange and Batey, 2006) validated the hypothesis that each of these RNAs uses a distinct mechanism for recognition of SAM. These studies provided insight into the functional groups of SAM that are required for specific ligand recognition in each case. In the S box and SAM-II riboswitches, nearly all of the functional groups of SAM make either direct or indirect contacts with residues in the RNA. Analog binding studies with these two riboswitches verified that modification of virtually any of the functional groups of SAM results in a dramatic loss of ligand binding affinity (Corbino et al., 2005; Lim et al., 2006; Winkler et al., 2003). One exception is in the case of the S box riboswitch, which was shown to be relatively tolerant of modifications at the methyl group, consistent with findings from the crystal structure that showed that this moiety is located
within a solvent-filled cavity in the ligand binding pocket of the RNA. The crystal structure of the SMK box riboswitch in complex with SAM revealed that the RNA forms extensive interactions with the adenosine and sulfonium moieties of the ligand. However, the methyl group attached to the SAM sulfonium and the amino acid side chain do not appear to be specifically recognized. Analog binding studies with the SMK box riboswitch will provide details about the stringency of recognition of this third class of SAM binding riboswitch, and allow further comparisons with the S box and SAM-II riboswitch systems.

In this study we employed a fluorescence based assay to study ligand binding to the SMK box riboswitch. We constructed a bipartite RNA that contains a 2-aminopurine (2-AP) substitution in the core of the riboswitch at a location that is predicted to undergo local conformational changes in response to SAM binding. We demonstrated that SAM addition results in a concentration-dependent quenching of 2-AP fluorescence, and that titration experiments can be used to obtain an apparent equilibrium binding affinity. This assay was used to investigate the effects of mutations in the RNA primary sequence on ligand binding, and also to test a variety of analogs for their ability to bind to the riboswitch. Our results verify that SAM analogs containing alterations at the methionine moiety and at the methyl group are tolerated for binding to the SMK box riboswitch. These data highlight differences between the ligand recognition requirements of this RNA and the other SAM binding riboswitches. However, our data support the model that all three SAM-responsive riboswitches utilize a similar strategy, which relies on the positive charge on the SAM sulfur group, to selectively recognize SAM over SAH.
We also used the 2-AP labeled RNA to measure the stability of the SAM-SMk
box complex in vitro. We showed that the time required for complex dissociation is
shorter than the half-life of the metK transcript inside the cell (A.M. Smith and R.T.
Fuchs, unpublished). We directly demonstrated that the conformational state of the metK
leader can be reversibly altered during multiple rounds through the direct modulation of
SAM levels in vitro. We conclude that the metK leader may be a reversible riboswitch
that is able to make multiple SAM-dependent regulatory decisions during the course of
its lifetime. This is in contrast to the B. subtilis yitJ S box element which is predicted to
be an irreversible riboswitch (Tomsic et al., 2008).

5.2 Materials and Methods

5.2.1 RNA constructs

RNA constructs corresponding to residues 15-118 (full-length transcript) or 47-
118 (downstream half of the bipartite RNA) relative to the predicted transcription start
site of the E. faecalis metK leader sequence were generated using T7 RNAP transcription.
Templates for in vitro transcription reactions were generated by ligating pairs of
complementary DNA oligonucleotides (purchased from Integrated DNA Technologies,
Inc., Coralville, IA) containing overlapping sequences as previously described (Yousef et
al., 2003). A promoter element recognized by T7 RNAP was positioned upstream of the
metK leader. Nucleotide substitutions in mutant constructs were incorporated by
replacing the complementary oligonucleotides encoding the corresponding nucleotides.
Ligated templates were amplified by PCR and sequences were verified by DNA
sequencing (Genewiz, Inc.). RNAs were synthesized using an AmpliScribe T7 High
Yield Transcription Kit (Epicentre Biotechnologies). Resulting products were purified by denaturing PAGE and eluted in 1X TBE followed by phenol:chloroform extraction and ethanol precipitation. Transcripts were resuspended in DEPC treated ddH2O and stored at -20 ºC. RNA corresponding to the upstream half of the bipartite metK leader (residues 13-46 relative to the predicted transcription start site), containing a 2-AP substitution at reside A29, was purchased from Dharmacon.

5.2.2 Fluorescence spectroscopy with SAM and related compounds

Bipartite RNA halves (1 µM each) corresponding to positions 13-46 and 47-118 of the E. faecalis metK leader sequence were combined in 1X transcription buffer (20 mM Tris-Cl, pH 8.0, 20 mM NaCl, 10 mM MgCl2, 0.1 mM EDTA) and heated to 65 ºC for five min followed by slow cooling to 22 ºC. For SAM binding experiments with the wild-type RNA, SAM (0.1-10 µM) was added and reactions were incubated at 22 ºC for 30 min prior to measuring 2-AP fluorescence. Experiments with U72C and G88A mutant RNAs, and with SAM analogs, were conducted identically except that the range of ligand concentrations was adjusted to facilitate determination of apparent $K_D$ values. Analog compounds were kindly provided by the laboratory of Dr. Squire Booker (Pennsylvania State University, University Park, PA). 2-AP fluorescence emission spectra (from 330-420 nm) were collected at 22 ºC using a FluoroLog-3 spectrofluorometer (Horiba Scientific) with an excitation wavelength of 310 nm. A maximum peak was observed at 375 nm and values at this wavelength were used to monitor fluorescence quenching in response to ligand addition. Fluorescence obtained with a sample containing 1X transcription buffer alone was subtracted to account for background. Percent quenching
in response to ligand was determined using the equation \( \Delta F/F_0 \times 100 \), where \( \Delta F \) represents the change in fluorescence relative to the initial fluorescence \( (F_0) \) observed in the absence of ligand. Percent quenching was plotted relative to ligand concentration and data were analyzed using non-linear regression analysis (KaleidaGraph, version 3.51). Reported \( K_D \) values represent the averages obtained from at least two independent experiments.

5.2.3 **Competition binding assays with SAM analogs**

RNA (1 µM) corresponding to positions 15-118 of the *metK* leader sequence was heated to 65 ºC in 1X transcription buffer for five minutes followed by slow cooling to 22 ºC. [Methyl-\(^3\)H]-SAM (1 µM, 15 Ci mmol\(^{-1}\); GE Healthcare) was added and reactions were incubated for 10 minutes at 22 ºC followed by addition of a competitor compound (100 µM) or buffer alone. Reactions were incubated for an additional 30 min at 22 ºC prior to being loaded onto 10 kDa MWCO Nanosep Centrifugal Devices (Pall Corporation). Samples were centrifuged at 14,000 X g for 35 sec and the relative amount of [methyl-\(^3\)H]-SAM retained on top of the filter was determined using scintillation counting. Percent retention was normalized to that observed with the negative control sample containing no competitor compound.

5.2.4 **In vivo translational fusions and \( \beta \)-galactosidase assays**

Studies to investigate the *in vivo* repression of a *metK-lacZ* translational fusion in response to SAM and methionine were performed as follows. *B. subtilis* strain BR151 (*metB10 lys-3 trpC2*) containing an *E. faecalis metK-lacZ* translational fusion
incorporated into the chromosome was constructed previously (Fuchs et al., 2006).
Briefly, the *E. faecalis* met*K* leader sequence (including the first fifteen nucleotides of the
coding region) was positioned downstream of a *B. subtilis* glyQS promoter, and was
fused to codon 18 of *lacZ* in plasmid pFG328 (Grundy and Henkin, 1993). The construct
was integrated into the *B. subtilis* chromosome using an SPβ prophage. *B. subtilis* strain
BR151 (*metB10 lys-3 trpC2*) containing a *yitJ-lacZ* transcriptional fusion was also
constructed previously using a similar method. Cells were grown in Spizizen minimal
media (Anagnostopoulos and Spizizen, 1961) supplemented with methionine (335 μM) to
early exponential phase. Cells were harvested and resuspended (at time zero; T₀) in
Spizizen minimal media with or without methionine (335 μM). Alternatively, cells were
grown in Spizizen minimal media without methionine for 120 min prior to addition of
methionine or SAM. Aliquots were collected every 30 min for determination of cell
density and β–galactosidase activity. Cell density was determined by measuring the
optical density at 595 nm, and β–galactosidase activity was measured after cells were
permeabilized with toluene as described (Miller, 1972). Studies to investigate the *in vivo*
repression of a met*K*-lacZ translational fusion in response to SAM analogs were
performed similarly except that analogs were added after 120 min of growth in the
absence of methionine and only a single time point was taken at 210 min (90 min after
addition of analog). Insertional disruption of the spe*D* gene in *B. subtilis* strain BR151
was carried out as previously described (Grundy and Henkin, 1998) and the 1A5 met*E*
mutation strain (*glyB133 metE tre-12 trpC2*) was obtained from the *Bacillus* Genetic
Stock Center (The Ohio State University, Columbus, OH). Growth curves and β-
galactosidase activity assays were performed at least in duplicate.
5.2.5 Measurement of SAM-SMK box complex half-life in vitro

Dissociation of SAM from the SMK box riboswitch was measured using stopped-flow spectroscopy. RNA halves (2 μM each) were annealed by combining in 1X transcription buffer and heating to 65 ºC for 5 min followed by slow cooling to 22 ºC. SAM (2 μM) was added and reactions were incubated at 22 ºC for 30 min to allow complex formation to reach equilibrium. A competitor RNA (10 or 20 μM), corresponding to positions 14-151 relative to the transcription start site of the B. subtilis yitJ leader RNA, was synthesized by T7 RNAP transcription and folded by heating to 65 ºC and slow cooling to 22 ºC as described above. Kinetic experiments were performed in an SX18-MV spectrometer (Applied Photophysics) using an excitation wavelength of 310 nm. Equal volumes of preformed SMK box-SAM complex and competitor yitJ leader RNA were rapidly mixed and fluorescence emission was monitored after passage through a 360 nm cutoff filter. For each experiment six individual traces were averaged and fit to either a single or double exponential equation using KaleidaGraph version 3.51 (Synergy Software) to determine the dissociation rate constant (k_{off}) and complex half-life.

5.2.6 Conformational switching of the SMK box riboswitch using fluorescence spectroscopy

Bipartite E. faecalis metK and B. subtilis yitJ leader RNAs were prepared and refolded as described above except that samples were slow cooled to 4 ºC instead of 22 ºC to slow down the kinetics of ligand association and dissociation. Emission fluorescence from the metK leader RNA was monitored at 375 nm (with an excitation wavelength of 310 nm) using the constant wavelength analysis application on a
FluoroLog-3 spectrofluorometer (Horiba Scientific). Fluorescence intensity was recorded at 2 sec intervals for a total of 420 sec, and either SAM or yitJ leader RNA (to sequester free SAM) was added at 60, 180, and 300 sec as indicated. Background fluorescence was determined using samples containing buffer alone (to which SAM and yitJ were added in a similar manner) and was subtracted from fluorescence intensity values obtained with samples containing 2-AP labeled metK leader RNA. Experiments were repeated in quadruplicate and data were plotted relative to time to monitor changes in fluorescence intensity in response to either addition or depletion of SAM.

5.3 Results

5.3.1 Design of a bipartite RNA containing a 2-AP substitution

To construct a fluorescent RNA corresponding to the E. faecalis metK leader sequence, we designed a bipartite molecule consisting of two RNA halves such that the upstream half was sufficiently small to facilitate efficient chemical synthesis (Dharmacon) and site-specific incorporation of a 2-AP probe (Fig. 5.1). The 34-nt upstream RNA used for these studies corresponds to residues 13-46 relative to the predicted transcription start site of the metK leader sequence, and the downstream RNA (synthesized by T7 RNAP transcription) corresponds to residues 47-118. This design allowed efficient hybridization between the two RNA pieces since the nick between the RNA halves lies at the top of the SmK box terminal loop.
Fig. 5.1. Bipartite RNA used to investigate ligand binding to the $S_{MK}$ box riboswitch. The bipartite RNA used for fluorescence studies is comprised of two RNA halves corresponding to residues 13-46 and 47-118. The nick between the two molecules is located at the $S_{MK}$ box riboswitch terminal loop. Residues 29-31 and 67-69, which base pair when SAM is bound, are shown in purple. The location of the 2-AP substitution at residue 29 is indicated with an arrow.
For placement of the 2-AP substitution, we considered regions in the RNA that become more structured upon SAM binding. Residues A29-G31 and C67-U69 exhibit decreased accessibility to a complementary oligonucleotide when ligand is bound (Fuchs et al., 2006; see section 3.3.2). Determination of the crystal structure of the SMK box riboswitch in complex with SAM confirmed that these residues form a portion of helix P2, which is one of three helical domains that converge in a three-way junction to form the SAM binding pocket (Lu et al., 2008; see section 3.3.4). Results from mutational analyses revealed that formation of these base pairs is essential for SAM binding (see section 3.3.4). Finally, results from NMR spectroscopy with a 59-nt minimal SMK box element demonstrated that imino signals from these core residues are detectable only in the presence of SAM, suggesting that base pairing occurs in response to SAM binding (see Chapter 4). These results confirm that residues A29-G31 and C67-U69 undergo local conformational changes in response to SAM binding and are good candidates for substitution with a fluorescent probe. 2-AP is a fluorescent purine analog that base pairs with uracil or thymine and is frequently used to probe local conformational changes in nucleic acid structure. Therefore, we selected residue A29 for substitution with a 2-AP probe to monitor structural alterations in the SMK box P2 helix in response to SAM binding.

5.3.2 Determination of SAM binding using spectrofluorimetry

To establish the suitability of the bipartite RNA for measuring changes in RNA conformation, we monitored emission fluorescence in a variety of reaction conditions. Fluorescence of the 2-AP labeled 34-nt upstream half of the bipartite RNA (1 µM) alone
was monitored from 330 to 420 nm and revealed a maximum emission peak at 375 nm (Fig. 5.2A). Fluorescence was partially quenched in samples in which the downstream RNA (1 µM) was added, probably due to hybridization between the two molecules. Addition of SAM (10 µM) to the bipartite RNA resulted in additional fluorescence quenching. This effect was not observed in samples containing the same concentration of SAH (data not shown), suggesting that fluorescence quenching is specific to SAM and is likely to represent structural rearrangements in the RNA brought about by specific binding. No significant fluorescence signal was detected in samples containing reaction buffer alone, or in samples containing reaction buffer plus SAM.
Fig. 5.2. SAM binding to the 2-AP labeled metK RNA. A) Emission fluorescence was monitored from 330-420 nm with an excitation wavelength of 310 nm. Maximal fluorescence was observed with the upstream RNA alone (US). Addition of the downstream RNA (US+DS) resulted in partial fluorescence quenching, and addition of SAM (US+DS+SAM) resulted in further quenching. B) Concentration-dependent quenching of fluorescence was observed with increasing concentrations of SAM (0.2-10 µM). C) Plot of percent quenching versus SAM concentration for the wild-type RNA (open diamonds). Nonlinear regression analysis was used to determine an apparent $K_D$ of 0.43 µM (average from three independent experiments). Fluorescence quenching of the U72C mutant RNA (closed diamonds) in response to SAM is shown for comparison. D) Fluorescence quenching was also observed with a bipartite metK leader sequence containing a G88A substitution, although binding affinity was decreased relative to wild-type. An apparent $K_D$ value for the G88A mutant construct was determined to be approximately 16 µM (~40-fold higher than wild-type) based on two independent experiments.
To determine whether fluorescence quenching in response to SAM is concentration-dependent, experiments were repeated with increasing SAM (0.1-10 μM) while the concentration of RNA was held constant (1 μM). Increased fluorescence quenching was observed with increasing SAM (Fig. 5.2B) and percent quenching was plotted relative to SAM concentration to determine the apparent binding affinity (Fig. 5.2C). These experiments yielded an approximate $K_D$ of 0.43 μM, which is in close agreement to the value (0.85 μM) that was determined using a size exclusion filtration assay (see section 3.3.3; Lu et al., 2008).

To validate the assay we selected two mutant metK leader RNAs that had previously been shown to exhibit decreased SAM binding activity relative to wild-type. A U72C mutation is predicted to disrupt favorable electrostatic interactions between the O4 carbonyl of residue U72 and the positive charge on the sulfonium of SAM, and was previously shown to result in a ~70-fold reduction in SAM binding activity using a size exclusion filtration assay (see section 3.3.5). No substantial change in fluorescence was observed with the U72C construct in response to SAM up to 100 μM (Fig. 5.2C and data not shown). These results confirm that the U72 residue is crucial for SAM binding to the SMK box riboswitch, and demonstrate that fluorescence quenching observed with the wild-type RNA is likely to be specific.

The G88A mutation, which disrupts base pairing in the P4 linker helix of the SMK box riboswitch, was previously shown to result in a ~75-fold reduction in binding activity in a size exclusion filtration assay (see section 3.3.5). The apparent $K_D$ for the G88A mutant was determined to be ~16 μM (Fig. 5.2D), which is approximately 40-fold higher than the wild-type value. In addition, the overall change in fluorescence (i.e., the
maximal quenching) in response to SAM was lower for the G88A mutant. This indicates that although the G88A construct retains the ability to bind SAM at high concentrations, the conformational change in response to ligand binding may not be identical to that observed with the wild-type RNA. These results provide additional evidence that the fluorescence change observed in response to SAM with the wild-type RNA is likely to be caused by specific binding, and demonstrate that the fluorescence assay described here is suitable for measuring the effects of nucleotide substitutions in the SMK box sequence on ligand recognition.

5.3.3 Analog binding to the SMK box RNA

To determine the specificity of ligand binding to the SMK box riboswitch, we tested a variety of SAM-related compounds for binding to the bipartite metK leader RNA using fluorescence spectroscopy (Fig. 5.3). Each of the analogs tested differs from SAM at a single functional group, with most of the compounds exhibiting alterations at the sulfur group or within the amino acid side chain. Titration experiments were performed and 2-AP fluorescence was plotted relative to ligand concentration to determine an apparent binding affinity for each of the related compounds (Fig. 5.3).
Fig. 5.3. SAM analog structures and apparent $K_D$ values determined by fluorescence spectroscopy. The chemical structure of each of the analogs tested is shown with modifications relative to SAM (top left) indicated by pink spheres. Numbers represent the average apparent $K_D$ values determined for each analog based on at least two independent experiments, except for 3’deoxy-SAM for which no binding curve could be obtained using the fluorescence assay (ND).

Se-SAM and Te-SAM differ from SAM only by the identity of the chemical element at the sulfur position. Se-SAM, which contains selenium in place of the sulfur, bound the metK leader with an apparent $K_D$ of 1.0 µM (2-fold higher than SAM). Te-SAM, which contains tellurium at the sulfur position, also bound the SMK box RNA, but with an apparent $K_D$ of 3.0 µM (7-fold higher than SAM). Sulfur, selenium, and tellurium are all chalcogen elements that exhibit many similar chemical properties, and so it is not
surprising that Se-SAM and Te-SAM retain some ability to bind to the $S_{MK}$ box RNA. Both Se-SAM and Te-SAM retain the positive charge at the central functional group, a feature that is predicted to be crucial for ligand recognition. Differences in the relative binding affinities of the two analogs are likely to be due to differences in the chemical properties of the atoms residing at the sulfur position, such as the Van Der Waals radii, which are larger in the two heavier chalcogens. It appears that as the Van Der Waals radius of the atom at the sulfur position increases ($S<Se<Te$) the binding affinity for the analog decreases ($S>Se>Te$).

To further investigate the requirement for the positive charge on the sulfur for binding to the $S_{MK}$ box riboswitch, we investigated binding of the naturally occurring analog SAH to the $metK$ leader RNA. It was previously demonstrated in a competition binding experiment that SAH supplied in a 50-fold molar excess over SAM is unable to compete for ligand binding to the $S_{MK}$ box riboswitch (Fuchs et al., 2006). Fluorescence spectroscopy revealed that SAH binds the $S_{MK}$ box with an apparent $K_D$ of 79 µM (Fig. 5.3), which represents a ~180-fold reduction in binding affinity relative to SAM. These results further validate the model that the positive charge on the sulfur group is essential for ligand recognition. The identity of the covalently attached methyl group does not appear to be critical, since substitution of this position with an ethyl group in the analog SAEt resulted in efficient ligand recognition and an apparent $K_D$ of 1.1 µM (~2.5-fold higher than for SAM). Also, the natural antibiotic sinefungin, which lacks the sulfonium ion and contains an amine substitution in place of the methyl group, retained the ability to bind the $metK$ leader, but with an affinity that is ~7-fold lower than that of SAM. The
decreased affinity for this analog is likely to be due to the decreased net positive charge density associated with the loss of the central sulfonium ion.

For the most part, SAM analogs containing alterations in the methionine tail retained the ability to bind to the SMK box element. For example, substitution of the α-amine of the methionine moiety by a hydroxyl in the analog OH-SAM did not significantly disrupt binding to the RNA (Fig. 5.3). Similarly, MP-SAM, which lacks both the α-amine and α-carboxyl groups of the amino acid, bound the metK leader with an apparent $K_D$ of 1.5 µM. Lastly, ME-SAM, which lacks these groups and is also missing the α-carbon of methionine, also efficiently bound the RNA with an apparent $K_D$ of 0.86 µM. These data support the hypothesis that the methionine tail is not specifically recognized by the SMK box riboswitch, consistent with the crystal structure (Lu et al., 2008). In contrast, SAC, in which the α-amine and α-carboxyl groups of the amino acid remain intact, but are brought close to the central sulfur moiety, exhibited a dramatic decrease (~170-fold) in binding to the metK leader RNA compared to SAM. The decrease in binding affinity may be due to steric clashes brought about by the closer proximity of these functional groups in the SAC structure.

Spectrofluorimetry experiments involving the analog 3’deoxy-SAM were inconclusive, as titration of this compound with the 2-AP labeled bipartite RNA yielded a nonspecific increase in fluorescence (data not shown). However, it is predicted based on findings from the crystal structure that alteration of the 3’hydroxyl group of SAM will result in a reduction in ligand binding, since this functional group makes key hydrogen bond interactions with residue G89 in the SD sequence of the riboswitch. This hypothesis
is supported by results obtained in a competition binding experiment using size exclusion filtration (see below).

### 5.3.4 Inhibition of SAM binding by related compounds using size exclusion filtration

To verify results obtained in the fluorescence based assay, we conducted competition binding experiments to assess the ability of the SAM-related compounds to prevent binding of SAM to the S_{MK} box riboswitch (Fig. 5.4). In these experiments, unlabeled SAM or SAM analogs (100 µM) were added to reactions containing [³H]-SAM (1 µM) incubated with \textit{metK} leader RNA (1 µM). Unbound SAM was separated by size exclusion filtration and radiolabeled SAM retained on top of the filter was quantitated. No significant amount of radiolabeled SAM was retained by the filter in the absence of RNA. Inclusion of \textit{metK} leader RNA in the reaction mix resulted in retention of [³H]-SAM, and retention was >10-fold decreased in the presence of excess unlabeled SAM (Fig. 5.4). Addition of Se-SAM, Te-SAM, SAEt, OH-SAM, MP-SAM, and ME-SAM resulted in a reduction of retention of [³H]-SAM similar to that observed with addition of SAM, indicative of efficient binding of the analog to the riboswitch RNA. Sinefungin and 3’deoxy SAM also resulted in reduced binding of [³H]-SAM, but to a lesser degree, indicating that these analogs bind the \textit{metK} leader RNA with decreased affinity relative to SAM. These results confirm that the 3’hydroxyl of the ribose group of SAM, which is missing from 3’deoxy SAM, is important for SAM binding, consistent with the crystal structure. In contrast, SAH and SAC did not efficiently compete for binding to the \textit{metK} leader, since the levels of [³H]-SAM retained in reactions containing these compounds
were similar to that observed in the absence of competitor. These results corroborate data obtained using fluorescence spectroscopy and validate the utility of the fluorescence system for studying determinants of ligand recognition in the S_{MK} box riboswitch.

![Graph showing competitive inhibition of SAM binding by related compounds.](image)

**Fig. 5.4. Competitive inhibition of SAM binding by related compounds.** *E. faecalis* metK leader RNA corresponding to positions 15-118 generated by T7 RNAP transcription was heated and refolded in the presence of [\(^3\)H]-SAM. Reactions were incubated with a 100-fold molar excess of unlabeled SAM analogs (indicated along the x-axis) and unbound radiolabeled-SAM was removed by size exclusion filtration. SAM binding is expressed as a percentage of that observed in the absence of competitor (none). neg, percent retention in the absence of metK leader RNA.

5.3.5 **Repression of S_{MK} box-regulated gene expression by SAM analogs in vivo**

It was previously shown that intracellular levels of SAM in *B. subtilis* strain BR151 (which is a methionine auxotroph) can be modulated through the addition or depletion of methionine in the growth media (Tomsic et al., 2008). This is due to the fact that the enzyme SAM synthetase (encoded by the essential metK gene) produces SAM from methionine and ATP. Expression of an *E. faecalis* metK-lacZ translational gene
fusion integrated in single copy into the chromosome of *B. subtilis* BR151 is repressed in response to methionine (Fuchs et al., 2006). However, the effect of addition of methionine to cells in which *metK-lacZ* expression is high due to methionine starvation was not known. We monitored *metK-lacZ* expression in *B. subtilis* cells grown in the presence or absence of methionine, and in cells to which methionine was added after 2 h of growth in medium lacking methionine. We also measured β-galactosidase activity in cells to which SAM was added (after 2 h of growth under inducing conditions) to determine whether exogenous SAM can act as a repressor of *metK-lacZ* expression *in vivo*. β-galactosidase activity was measured at T₀ and every 30 min thereafter for a total duration of 210 min.

In cells grown in the presence of methionine (added at T₀), low levels of β-galactosidase activity were observed throughout the course of the experiment, consistent with previous results (Fig. 5.5A; Fuchs et al., 2006). In cells to which no methionine was added, initial levels of β-galactosidase activity were also low, presumably due to residual methionine in the resuspended cells. β-Galactosidase activity started to increase after 90 min of growth in the absence of methionine, and after 210 min *metK-lacZ* expression was derepressed by ~4-fold. Addition of either methionine or SAM (at 120 min) was sufficient to reverse the induction phenotype, and after 210 min β-galactosidase activity had returned to near basal levels. Some of the remaining β-galactosidase activity is likely to be due to residual β-galactosidase inside the cells and not to newly synthesized protein.
Fig. 5.5. Repression of lacZ fusions in B. subtilis upon addition of methionine or SAM. (A) Translational E. faecalis metK-lacZ fusion. (B) Transcriptional B. subtilis yitJ-lacZ fusion. Cells were grown in Spizizen medium (Anagnostopoulos and Spizizen, 1961) until early log phase and collected by centrifugation. Cells were resuspended in fresh medium with or without methionine (335 µM). After 120 min, methionine or SAM (335 µM) was added to cultures as indicated. Values represent the averages from at least two independent experiments. MU, Miller units (Miller, 1972).

Similar experiments were conducted with B. subtilis cells containing a yitJ-lacZ transcriptional fusion. The yitJ leader sequence from B. subtilis contains an S box riboswitch that regulates expression via transcription attenuation in response to SAM
binding. Consistent with previous results (Fuchs et al., 2006; Grundy and Henkin, 1998), the stringency of regulation was higher for the \textit{yitJ-lacZ} transcriptional fusion, as \( \beta \)-galactosidase activity was nearly undetectable in cells grown in the presence of methionine (Fig. 5.5B). After 90 min of growth, \( \beta \)-galactosidase activity started to increase in cells lacking methionine, and after 210 min of growth \textit{yitJ-lacZ} expression was derepressed by \( \sim160 \)-fold. As with the \textit{metK-lacZ} fusion, addition of methionine or SAM to cells containing the \textit{yitJ-lacZ} fusion (after 120 min of growth in media lacking methionine) resulted in rapid repression of \( \beta \)-galactosidase expression. Taken together, these results suggest that addition of methionine or SAM to the growth media can stimulate rapid repression by SAM-responsive riboswitches \textit{in vivo}. These results indicate that SAM added to the growth media is likely to be imported into the cell where it acts as a repressor of \textit{metK-lacZ} expression. Alternatively, it is possible that SAM is converted to methionine (via a metabolic pathway for methionine biosynthesis or through spontaneous hydrolysis), which can then be used as a substrate for SAM synthesis. To further investigate these two possibilities, we utilized \textit{B. subtilis} mutant strains containing disruptions in genes that are involved in SAM recycling and methionine biosynthesis.

The first \textit{B. subtilis} strain that we tested (BR151-SpeDKO) was one in which the \textit{speD} gene, which encodes SAM decarboxylase, has been inactivated. SAM decarboxylation constitutes one of the early steps in the methylthioadenosine (MTA) recycling pathway that ultimately leads to the formation of methionine from 2-keto-4-methylthiobutyrate (Fig. 5.6; Murphy et al., 2002). Therefore, disruption of the \textit{speD} gene is expected to block this pathway of methionine biosynthesis. A \textit{metK-lacZ} translational fusion was integrated in single copy into the chromosome of \textit{B. subtilis}
strain BR151-SpeDKO and β-galactosidase activity was monitored in cells grown in the presence or absence of methionine or SAM.

Results obtained with the \textit{B. subtilis} BR151-SpeDKO strain were similar to those observed with strain BR151. Expression of the \textit{metK-lacZ} translational fusion was high during growth in the absence of methionine, and was repressed by ~5-fold when
either methionine or SAM was added to the media (Fig. 5.7, left panel). These results indicate that addition of either of these compounds can repress the \textit{metK-lacZ} fusion. The observation that SAM acts at least as effectively as methionine, and with similar kinetics, suggests that exogenous SAM is acting directly rather than through conversion to methionine and back to SAM.

![Fig. 5.7. Expression of \textit{metK-lacZ} translational fusion and growth of cells in the presence or absence of SAM and related compounds. The \textit{metK-lacZ} translational fusion was integrated into the chromosome of \textit{B. subtilis} strain BR151 SpeDKO (gray bars) or 1A5 SpeDKO MetEKO (black bars). Cells were grown at 37 °C in Spizizen minimal media (Anagnostopoulos and Spizizen, 1961) containing methionine until early exponential phase. Cells were harvested and resuspended in Spizizen minimal media (lacking methionine), and methionine, SAM, or SAM analogs were added to cells after 120 min of growth as indicated. Samples were taken after an additional 90 min of growth and used for determination of β-galactosidase activity (left panel) and cell density (right panel). MU, Miller units (Miller, 1972).](image-url)
We also tested several SAM analogs for their ability to act as repressors of \textit{metK-lacZ} expression in the \textit{B. subtilis} BR151-SpeDKO strain (Fig. 5.7, left panel). Addition of Se-SAM and ME-SAM was sufficient to repress \textit{metK-lacZ} expression by 5-fold and 2-fold respectively, consistent with the \textit{in vitro} fluorescence assay data, which showed that these SAM analogs retain the ability to bind to the \textit{metK} leader sequence (see sections 5.3.3). In contrast, addition of sinefungin or SAH to the growth media did not result in repression of \(\beta\)-galactosidase activity. These data are also consistent with the \textit{in vitro} experiments, which showed that these two analogs bind the \textit{metK} leader with decreased affinity relative to SAM (see sections 5.3.3).

Typically, starvation for methionine leads to an inhibition of growth (relative to cells grown in the presence of methionine) of \textit{B. subtilis} strain BR151, since these cells are auxotrophic for the amino acid (see Fig. 5.7, right panel, -Met). This phenotype could be caused by decreased translation in response to low levels of methionine. However, the observation that \(\beta\)-galactosidase expression continues to increase in cells that exhibit inhibited growth in the absence of methionine suggests that methionine levels are sufficient to support translation. Therefore, the growth phenotype observed in response to methionine starvation may be attributable to something besides a shortage of the amino acid. One possibility is that limited methionine inside the cell may be preferentially used for translation, resulting in a shortage of SAM. Reduced levels of SAM would affect methyltransferase reactions inside the cell and also cause decreased cell growth. This would explain why addition of SAM or Se-SAM to the media results in restoration of cell growth and rapid repression of the \textit{metK-lacZ} fusion (Fig. 5.7). Addition of ME-SAM to the growth media, which resulted in partial repression (~2-fold) of \textit{metK-lacZ} expression,
was not sufficient to restore growth of methionine starved cells (Fig. 5.7). This suggests that the cells are unable to utilize ME-SAM (which lacks the amine and carboxyl groups as well as the α-carbon of the methionine moiety of SAM) to accommodate a shortage of SAM or methionine.

To further investigate the effect of SAM analogs on the in vivo expression of the metK-lacZ translational fusion, we utilized a B. subtilis strain in which two genes involved in methionine biosynthesis have been disrupted. This strain (B. subtilis 1A5-SpeDKO-MetEKO) contains the speD gene disruption described above, in addition to a mutation in the metE gene, which is responsible for the conversion of homocysteine to methionine in the primary pathway of methionine biosynthesis (Fig. 5.6; Murphy et al., 2002). This mutation differs from the metB mutation present in strain BR151 in that it prevents recycling of SAH to regenerate methionine and SAM. The metK-lacZ translational fusion was integrated into the chromosome in single copy, and β-galactosidase activity was measured in cells grown in the presence or absence of methionine, SAM, and SAM analogs.

The pattern of repression observed with the 1A5-SpeDKO-MetEKO strain was similar to that observed with the single gene disruption strain BR151-SpeDKO (Fig. 5.7). β-Galactosidase activity was high in cells grown in the absence of methionine, and addition of methionine, SAM, Se-SAM and to a lesser degree ME-SAM resulted in repression of metK-lacZ expression, while addition of sinefungin or SAH did not. Restoration of growth was again observed with SAM and Se-SAM suggesting that addition of these compounds may replenish SAM inside the cell. Taken together, these results suggest that SAM, Se-SAM, and ME-SAM added to the growth media are likely
to be imported into the cell where these compounds acts as a repressors of *metK-lacZ* expression.

5.3.6 **Measurement of the stability of the SAM-S\textsubscript{MK} box complex in vitro.**

Measurement of stability of the RNA-SAM complex provides insight into the potential reversibility of the S\textsubscript{MK} box riboswitch. If the t\textsubscript{1/2} of the complex is shorter than the t\textsubscript{1/2} of the mRNA, the riboswitch may have the opportunity to make multiple regulatory decisions during the course of its lifetime. Results from our laboratory revealed that the t\textsubscript{1/2} of the *metK* transcript in *E. faecalis* is \(~3\) min in cells grown in conditions of high methionine and \(~4.4\) min in cells grown in low methionine (R.T. Fuchs, unpublished). The increase in transcript stability when SAM levels are low is likely to be the result of the stabilizing effect of the translating ribosome when *metK* expression is derepressed (i.e., when the riboswitch TIR is available for translation initiation).

The rate of ligand dissociation was measured by rapidly mixing the equilibrated bipartite RNA-SAM complex with an excess of a SAM binding competitor RNA. The competitor RNA used was the *B. subtilis* yit\textsubscript{J} S box leader RNA, which binds SAM with an apparent *K*\textsubscript{D} of 0.019 \(\mu\text{M}\) (Tomsic et al., 2008). In this experiment, as SAM dissociates from the S\textsubscript{MK} box RNA it is sequestered by the yit\textsubscript{J} leader RNA, so that the rate of the change in fluorescence represents the *k*\textsubscript{off}. Fluorescence change was monitored in a stopped-flow spectrometer. Rapid mixing of the SAM-S\textsubscript{MK} complex with a 5-fold molar excess of competitor RNA resulted in an increase of 2-AP fluorescence emission (Fig. 5.8). At least 6 individual traces were averaged and fit to a single exponential
function to determine a $k_{\text{off}}$ of 0.089 sec$^{-1}$. Inspection of the residual plot revealed that a single exponential function provides a reasonable fit, and no improvement was observed when traces were fit to a double exponential function (data not shown).

Fig. 5.8. Dissociation of the SAM-S$_{\text{MK}}$ box riboswitch complex \textit{in vitro}. The dissociation of SAM from the S$_{\text{MK}}$ box riboswitch was monitored at 22 °C in a stopped-flow spectrometer. Syringe 1 was loaded with \textit{E. faecalis} S$_{\text{MK}}$ box RNA, containing a 2-AP fluorescent probe, in complex with SAM. Syringe 2 contained a 5-fold molar excess of a competitor RNA (\textit{B. subtilis} yitJ leader) that exhibits a high affinity for SAM. SAM dissociation resulted in an increase in 2-AP fluorescence over time. Data points (blue spheres) represent the averages from six individual traces and were fit to a single-exponential function (orange trace) to yield a dissociation rate constant of 0.089 sec$^{-1}$ and a t$_{1/2}$ of 7.8 sec. The residuals plot (bottom panel) shows the difference between the actual (blue spheres) and calculated (orange trace) data sets.
In order to verify that the observed rate directly reflects the rate of SAM dissociation, we repeated the experiment with a 10-fold molar excess of competitor RNA, and the rate was shown to be the same. From these results, the $t_{1/2}$ of the SAM-SM$_{MK}$ box complex was determined to be 7.8 sec \textit{in vitro}. The $t_{1/2}$ of the complex is \sim 20-fold shorter than the $t_{1/2}$ of the \textit{metK} transcript \textit{in vivo}. This suggests that the transcript is available for multiple rounds of SAM binding and dissociation, and that the SM$_{MK}$ box riboswitch is likely to be a reversible riboswitch that can potentially undergo several regulatory decisions within the lifetime of the mRNA in response to fluctuations in SAM availability.

\textbf{5.3.7 Reversibility of the SM$_{MK}$ box riboswitch \textit{in vitro}}

We utilized a time resolved fluorescence assay using the 2-AP labeled \textit{E. faecalis metK} leader RNA to visualize the ability of the SM$_{MK}$ box riboswitch to switch between alternate conformations in response to fluctuating levels of SAM. Fluorescence at 375 nm was monitored using constant wavelength analysis, with readings collected every 2 sec for a total of 420 sec. During the course of the experiment, SAM levels were modulated through the direct addition of SAM, or of a SAM sequestering molecule (the \textit{B. subtilis yitJ} leader RNA) at precise time intervals. Emission fluorescence from the 2-AP labeled bipartite RNA (1 \textmu M) in the absence of additional reaction components was monitored from 0-60 sec (Fig. 5.9). Under these conditions, a non-specific decrease in fluorescence was observed, which is likely to be due to photobleaching of the 2-AP fluorophore. Addition of SAM (2 \textmu M) at 60 sec resulted in a characteristic decrease in fluorescence intensity over time, consistent with the SAM-dependent quenching of
fluorescence reported previously (see section 5.3.2). The subsequent addition of *B. subtilis* *yitJ* leader RNA (4 µM), which acts as a SAM binding competitor molecule, resulted in a reversal of the quenching phenotype and an increase in fluorescence intensity over time. These data are consistent with results obtained in the SAM-*metK* leader RNA dissociation experiments (see Fig. 5.8 and section 5.3.6) and suggest that the SMK box riboswitch returns to a conformational state that resembles the ligand-free conformation as SAM levels are depleted. Lastly, addition of SAM (20 µM) in a 5-fold molar excess over the *yitJ* RNA at 300 sec resulted in another reversal in fluorescence intensity, as emission levels were once again quenched in response to ligand binding. These data directly demonstrate that the SMK box riboswitch can undergo multiple conformational switches in response to changing levels of SAM *in vitro*.
Fig. 5.9. Conformational switching of the SmK box in response to SAM using fluorescence spectroscopy. Emission fluorescence of an *E. faecalis metK* leader RNA (1 µM) containing a 2-AP fluorescent probe was monitored over time. Addition of SAM (2 µM) at 60 sec resulted in fluorescence quenching, while addition of *yitJ* leader RNA (4 µM) at 180 sec restored the fluorescence intensity. Addition of more SAM (20 µM) at 300 sec reversed the phenotype again, leading to an even greater decrease in fluorescence intensity over time. Gaps in the data represent the time required for addition and mixing of sequential reaction components. Values plotted represent the averages from four independent experiments.

5.4 Discussion

In the current study we developed a fluorescence based assay for monitoring conformational changes in the core of the SmK box riboswitch in response to ligand binding. We used this assay to determine an apparent SAM binding affinity of ~0.43 µM for the *E. faecalis metK* leader. This value is in close agreement with the value (~0.85 µM) obtained using a size-exclusion filtration assay. The fluorescence assay has the advantage that the concentration of reaction components remains unaltered throughout the course of the experiment, allowing for a more accurate determination of binding
affinity. We demonstrated that this assay can be used to investigate the effects of nucleotide substitutions in the riboswitch sequence on SAM binding, as constructs containing mutations at residues known to be important for ligand recognition (based on the crystal structure and size-exclusion filtration assays) resulted in reduced affinity for SAM.

We investigated the specificity of the SMK box riboswitch by testing a variety of SAM analogs for their ability to bind to the metK leader. Results from these experiments validated observations made from the crystal structure of a minimal SMK box element (see Chapter 3). For the most part, alteration of the methyl group or the methionine moiety of SAM resulted in a modest reduction (<3-fold) in binding affinity. However, alteration of the positive charge on the sulfur resulted in a dramatic decrease in binding affinity (~180-fold). These results provide additional evidence that the positive charge on the sulfur is essential for stabilizing favorable electrostatic interactions with the riboswitch RNA, a feature that is shared by all three of the SAM binding riboswitches. Deletion of the 3’hydroxyl of SAM also resulted in decreased binding affinity, consistent with the observation that this functional group makes hydrogen bond interactions with residue G89 of the riboswitch RNA.

Results from the analog binding assays revealed that the SMK box riboswitch is more tolerant of changes to the methionine and methyl moieties of SAM than was observed for the S box and SAM-II riboswitches. These results provide further evidence that the three riboswitches have evolved separately and exhibit unique ligand recognition properties. However, all three riboswitch classes share the ability to discriminate against the physiologically relevant analog, SAH, and utilize a common mechanism for doing so.
Each of these riboswitch systems has evolved to select for the positive charge on the sulfur, the primary feature that is missing from SAH. This common property of the SAM binding riboswitches ensures that SAH, the metabolic byproduct of SAM-mediated methyltransferase reactions, does not repress expression of \( \text{metK} \).

In addition to measuring the binding properties of the \( S_{\text{MK}} \) box riboswitch at equilibrium, the fluorescence assay allowed us to investigate the kinetics of ligand dissociation. A major question in the riboswitch field is whether metabolite binding riboswitches can act as reversible switches, rather than as sensors that monitor their cognate signal and make a single gene expression decision. We previously showed that the SAM-responsive \( S \) box riboswitch in \( B. \text{subtilis} \), which controls genes involved in biosynthesis of methionine and SAM, is likely to be irreversible. The \( S \) box RNA-SAM complex is highly stable, with a \( t_{1/2} \) that is in excess of 4 min (for most \( B. \text{subtilis} \) \( S \) box elements; Tomsic et al., 2008). The majority of \( S \) box riboswitches, including those in \( B. \text{subtilis} \), regulate gene expression at the level of premature termination of expression. This results in a short window of opportunity for SAM binding (and potentially for dissociation and rebinding) during transcription of the \( ~200 \)-nt leader RNA. These data indicate that the \( S \) box riboswitch is likely to be irreversible.

In this study, we have shown that the \( S_{\text{MK}} \) box riboswitch is likely to be a reversible riboswitch. We demonstrated switching of the \( \text{metK} \) leader between alternate conformations \( \text{in vitro} \) in response to fluctuating levels of SAM. \( \text{In vivo} \), the \( t_{1/2} \) of the \( E. \text{faecalis} \) \( \text{metK} \) transcript was found to be 2.8 min in cells grown under conditions of high methionine (R.T. Fuchs, unpublished), which is much greater than the half-life of the SAM-\( S_{\text{MK}} \) box complex (7.8 sec) \( \text{in vitro} \). This indicates that subsequent to complex
formation, SAM dissociation can occur before the mRNA is degraded, leaving the \textit{metK} transcript accessible for either translation initiation or interaction with another SAM molecule. The relative rate constants for each of these processes, and the concentrations of reaction components available at a given time, determine which of these events is more likely to occur. A reversible mechanism of regulation at the level of translation initiation ensures that cells are poised to utilize existing mRNA transcripts to respond rapidly to transient fluctuations in SAM pools. This mechanism would allow the cell to upregulate expression of the \textit{metK} gene quickly in response to decreased levels of SAM to ensure that SAM, which is the primary methyl donor for many essential metabolic processes, is available in sufficient quantities at all times.
CHAPTER 6

SUMMARY AND FUTURE DIRECTIONS

The work presented here focuses on two metabolite binding riboswitches (the THI box and SMK box) that regulate gene expression at the level of translation initiation. In both of these regulatory systems, ligand binding results in a structural rearrangement of the riboswitch RNA that leads to occlusion of the SD sequence and inhibition of translation of the downstream coding region. Despite these similarities, each of these RNA elements utilizes a different mechanism for ligand binding and regulatory control.

The THI box riboswitch, which binds TPP, is one of the earliest discovered riboswitches. It exhibits a modular architecture in which the ligand binding domain is separate from the regulatory domain. This modularity allows the ligand binding domain, which is characterized by conserved structural features, to be coupled to a variety of regulatory platforms. The THI box element located in the E. coli thiM gene regulates at the level of translation initiation, by sequestration of the SD sequence by a complementary ASD sequence; however, some THI box riboswitches utilize other regulatory mechanisms that affect transcription termination, alternative splicing, or mRNA stability (Miranda-Rios, 2007).
Based on phylogenetic analysis it was predicted that in the absence of TPP the *E. coli* thiM leader RNA favors a conformation in which the SD sequence is accessible for translation initiation (Miranda-Rios et al., 2001). In this conformation, the ASD sequence is sequestered by a complementary AASD sequence. Binding of TPP results in sequestration of the AASD sequence in a helix (P1) at the base of the ligand binding domain. In this arrangement, the ASD sequence is free to pair with the SD sequence and gene expression is downregulated. Here we show that TPP binding to the thiM leader RNA leads to decreased accessibility of the SD sequence and decreased binding of 30S ribosomal subunits, consistent with the model (see Chapter 2; Ontiveros-Palacios et al., 2008).

We performed mutational analyses of conserved residues in the ligand binding domain of the THI box riboswitch and identified mutations (Class I) that result in decreased TPP binding activity and a subsequent loss of TPP-dependent repression (see Chapter 2; Ontiveros-Palacios et al., 2008). The SD sequences in Class I mutant constructs are accessible regardless of the presence or absence of TPP. We also identified mutations (Class II) that result in occlusion of the SD sequence whether or not TPP is present. Constructs containing these mutations appear to have lost the ability to form the unliganded conformation of the THI box riboswitch. This is the first example of riboswitch mutations located in the ligand binding domain that lead to constitutive repression of gene expression.

The effect of the Class I mutations can be explained because of the loss of the ability of these RNAs to bind ligand. However, it is less clear how the Class II mutations result in SD sequence occlusion even in the absence of TPP. Additional structural studies
would provide insight into the effects of these mutations on the structural state of the RNA. X-ray crystallography of the ligand binding domain of a Class II mutant in the absence of TPP would reveal whether the ligand binding pocket resembles that of the wild-type construct, and also whether formation of the P1 helix at the base of the ligand binding domain is maintained. However, it may be difficult to obtain refracting crystals of a Class II mutant construct. In this case, alternative methods for investigating structure could be utilized such as enzymatic and chemical probing. These experiments would reveal whether Class II mutant constructs assume a conformation that resembles the ligand-bound state of the wild-type RNA, or whether they form a new conformation in which the SD sequence is occluded (even in the absence of ligand). Results from these studies would provide insight into the roles of individual nucleotides in stabilization of alternate conformations of the THI box riboswitch and into the interplay between the ligand binding domain and the regulatory platform.

The SMK box riboswitch is one of three SAM binding riboswitches that have been discovered (Corbino et al., 2005; Fuchs et al., 2006; McDaniel et al., 2003). Characterization of each of these riboswitch classes allows for the direct comparison of three RNAs that have evolved separately to recognize the same ligand. The SMK box and SAM-II riboswitches exhibit a simple single-domain architecture that has not been observed in other riboswitch classes. This compact design facilitates high-resolution structural investigations using the complete riboswitch element including both the ligand binding domain and the regulatory platform.

We have performed mutational analysis to define a minimal construct (based on the SMK box riboswitch from *E. faecalis*) that retains the ability to bind SAM (see
Chapter 3; Lu et al., 2008). We used size-exclusion filtration to demonstrate that a 51-nt construct binds SAM with a slightly higher affinity than the 106-nt full-length construct. A similar 53-nucleotide minimal construct was used to obtain the crystal structure of the SMK box riboswitch in complex with SAM (see Chapter 3; Lu et al., 2008). These studies confirmed that the SMK box riboswitch exhibits structural features in the ligand binding domain that are distinct from those of the other two SAM binding riboswitches. While the S box and the SAM-II riboswitches recognize virtually every functional group of SAM (except for the methyl group of SAM in the case of the S box riboswitch), the SMK box riboswitch does not specifically recognize the methyl or methionine groups of the ligand. The SMK box riboswitch makes extensive hydrogen bond interactions with the adenine moiety and ribose moieties of SAM. Also, favorable electrostatic interactions between a uracil residue in the RNA and the positive charge on the sulfonium of SAM are essential for ligand binding. This latter feature explains how the riboswitch discriminates against SAH and is the same mechanism used by the S box and SAM-II riboswitches to specifically recognize SAM.

Results from the crystal structure also provide a structural basis for the mechanism of regulation by the SMK box riboswitch. These studies showed that residues that comprise the SD sequence are required for stabilization of the ligand binding pocket, and that certain residues within the SD sequence make direct contacts (i.e., hydrogen bonding, electrostatic interactions, and base stacking) with SAM. For example, residue G90 (which corresponds to the central position of the SD sequence) is involved in base stacking with the adenine moiety of SAM and also participates in formation of a base triple (A73·G90·C25) that forms the floor of the SAM binding pocket. This explains why
the SD sequence of all known SMK box riboswitches is GGGGG instead of the consensus GGAGG.

We also investigated the conformation of the SMK box riboswitch in the absence of ligand. Based on phylogenetic analysis, it was predicted that in the absence of SAM, helices P1, P2, and P4 (which constitute the ligand binding domain) are missing, and instead a new helix (P0) is formed (Fuchs et al., 2006). We used NMR spectroscopy to verify this hypothesis and to demonstrate that residues at the 5’ end of the riboswitch element (that are not required for SAM binding) are required for stabilizing the unliganded conformation (see Chapter 4). These results provide the first high-resolution structural evidence for the unliganded conformation of a riboswitch using an RNA element that encompasses both the ligand binding domain and the regulatory platform. We also used NMR spectroscopy to show that the conformation of the SMK box riboswitch in solution (in the presence of SAM) is the same as that determined using x-ray crystallography. While the results from the NMR experiments provide evidence for the conformational switching of the SMK box riboswitch in response to SAM, and validate the secondary structural model, additional experiments are needed to determine the complete 3D structure of this riboswitch in the absence of ligand.

Another focus of my research has been to investigate the properties of ligand binding to the SMK box riboswitch. For these studies, we developed a fluorescence assay using a bipartite RNA containing a 2-AP probe in the core of the riboswitch (see Chapter 5). This assay was used to measure the binding affinity of the SMK box riboswitch for SAM and a variety of SAM analogs. Results from these experiments corroborated findings from the crystal structural and showed that modifications to the methyl group
and methionine tail of SAM do not inhibit ligand recognition. We also showed that the
$S_{MK}$ box riboswitch exhibits a ~180-fold preference for SAM over SAH, consistent with
the hypothesis that the positive charge on the sulfonium group of SAM is essential for
ligand recognition. Additional experiments are required to determine the effects of
modifications of the adenine moiety of SAM. It is predicted based on the crystal structure
that modification of these functional groups will result in a reduction of binding affinity,
since they are known to form hydrogen bonds with residues in the RNA.

We have also been interested in determining whether SAM analogs that retain
the ability to bind the $S_{MK}$ box riboswitch \textit{in vitro} can act as repressors of gene
expression \textit{in vivo}. We utilized a \textit{B. subtilis} strain containing a \textit{metK-lacZ} translational
fusion to test whether SAM, Se-SAM, and ME-SAM (all of which bind SAM \textit{in vitro})
can affect $S_{MK}$ box-dependent regulation (see Chapter 5). Addition of each of these
compounds to the growth media of methionine starved cells resulted in reduced $\beta$–
galactosidase activity suggesting that they may be acting as repressors. However, it is
possible that these compounds are being hydrolyzed to produce methionine, which can
then be converted by the cells to SAM. To circumvent this possibility we tested
repression of the \textit{metK-lacZ} translational fusion in \textit{B. subtilis} strains containing
disruptions in genes encoding key enzymes required for methionine biosynthesis and
recycling. These experiments yielded similar results, and addition of SAM, Se-SAM, and
ME-SAM to the growth media resulted in repression of \textit{metK-lacZ} expression. These
results suggest that these compounds may act as repressors of $S_{MK}$ box genes \textit{in vivo}.
Additional experiments are needed to determine whether other SAM analogs can act as
repressors of $S_{MK}$ box genes \textit{in vivo}. 

168
We have also investigated the rate of dissociation of the SMK box-SAM complex (see Chapter 5). These experiments revealed that the half-life of the complex is shorter than the half-life of the mRNA inside the cell. This indicates that the SMK box riboswitch may be a reversible riboswitch that makes multiple regulatory decisions during the course of its lifetime. Reversibility of the SMK box-SAM complex ensures that the metK transcript is poised for translation initiation as soon as SAM levels drop below the threshold required for repression. This is in contrast to the B. subtilis yitJ S box riboswitch, which appears to be an irreversible riboswitch (Tomsic et al., 2008). In addition to measuring the rate of complex dissociation, we are also interested in investigating the kinetics of ligand association. Experiments to characterize the pre-steady state kinetics of SAM association to the SMK box riboswitch could provide insight into the mechanism of ligand binding (i.e., whether it occurs as a one- or two-step process). Stopped-flow spectroscopy or surface plasmon resonance spectroscopy might provide a suitable methodology for acquiring kinetic parameters.

In summary, we have used a variety of methods to study the ligand binding properties and regulatory mechanisms of two different translational riboswitches. We have shown that point mutations in the THI box ligand binding domain can have dramatic effects on riboswitch function. We have shown that the SMK box riboswitch is poised for translation initiation and that SAM-dependent repression is likely to be a reversible process. This rapid response mechanism ensures that SAM, the primary methyl donor inside the cell, is always available in sufficient quantities.
REFERENCES


oryzae thia requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR. FEBS Lett 555, 516-520.


