2011

Structural and mechanistic studies on the interactions between S-adenosyl methionine (SAM) and the SAM-I riboswitch

Wei Huang
Louisiana State University and Agricultural and Mechanical College

Follow this and additional works at: https://digitalcommons.lsu.edu/gradschool_dissertations

Recommended Citation
Huang, Wei, "Structural and mechanistic studies on the interactions between S-adenosyl methionine (SAM) and the SAM-I riboswitch" (2011). LSU Doctoral Dissertations. 3535.
https://digitalcommons.lsu.edu/gradschool_dissertations/3535

This Dissertation is brought to you for free and open access by the Graduate School at LSU Digital Commons. It has been accepted for inclusion in LSU Doctoral Dissertations by an authorized graduate school editor of LSU Digital Commons. For more information, please contact gradetd@lsu.edu.
STRUCTURAL AND MECHANISTIC STUDIES ON
THE INTERACTIONS BETWEEN S-ADENOSYL METHIONINE (SAM)
AND THE SAM-I RIBOSWITCH

A Dissertation

Submitted to the Graduate Faculty of the
Louisiana State University and
Agricultural and Mechanical College
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

In

The Department of Biological Sciences

by

Wei Huang
B.E., Xi’an Jiaotong University, 2006
December, 2011
DEDICATION

This dissertation is dedicated to my grandfather, Qiyou Huang, and my uncle, Jianzhong Huang, for their endless supports for me to pursue further education. Also to my parents, Jianhua Huang and Chunjin Zheng, and my parents-in-law, Youzhong Xie and Jinping Zhu, for their open-minded on my career choice and their excellent role modes on work ethics. To my undergraduate advisor, Yili Wang, for his encouragement and supports for me to step into the field of bioinformatics and computational biology.

Most importantly, I dedicate this work to my beloved wife, soul mate and best friend—Lin Xie, who gave me the gift of loving her; for rescuing me from a colorless college life; for never allowing me to lose faith or giving up; for celebrating every little achievement that I’ve made; for supporting me in every way; for wanting the very best for me; for giving birth to our lovely daughter, Claire Huang, who made my world outside of research filled with joy; and for loving me with all of her heart.
ACKNOWLEDGEMENTS

First and foremost, I thank Dr. Fareed Aboul-ela for being my mentor and teacher for the past four years. Although my undergraduate GPA doesn’t fulfill the minimum requirement, he was brave enough to take me on as a graduate student. I would not have been able to work on an interdisciplinary research project without his vision and help. His instruction and advice have had a profound influence on me, and I am incredibly grateful. Next to Dr. Aboul-ela, the person teaches me the essentials of computational science, provides me working environment to nourish my computational skills, and supplies us with large amount of computational time on High Performance Computers, is Dr. Shantenu Jha. His philosophy on “being lazy” motives me to write better scripts and keep enlarging my arsenal of computational techniques. I am grateful to have both of them as my primary research support.

My dissertation committee of Dr. John R. Battista, Dr. Yong-hwan Lee and Dr. Rodrigo Valverde took their time to contribute many helpful ideas and provide much needed expertise. I also thank Dr. Battista for leading me into the intriguing microbial world and discussions on high-throughput sequencing technologies, and Dr. Lee for his insightful discussions on structure based drug design.

I would like to thank Dr. Marcia Newcomer, Dr. Sue Bartlett, Dr. Anne Grove, Dr. Grover Waldrop, Dr. Erin Hawkins and Dr. Simon Chang for lending their expertise to me whenever I had a question and providing access to materials and equipment in their laboratories. I would also thank Dr. Joohyun Kim at CCT for his helps and discussions on various topics related to our research projects. Also thanks to Dr. Thomas Weldeghiorghis and Dr. Dale
Treleaven for their assistance with using the NMR facility. I am also very grateful to the IT staffs, especially Dr. Le Yan, for their help with using High Performance Computers. A special thanks to our collaborator, Dr. Sichun Yang, for his help with collecting SAXS data, and to Dr. Marc Parisien, who taught me to become a master on MC-sym and provided insightful discussions on RNA folding.

I am grateful to all the current and former members in the Aboul-ela lab that have helped me in my research: Vamsi Boyapati, Patrick Robichaux, Sunayana Mitra, Gary LeBlanc, Sohyun Park, Shiqiu Li, Jessica Spedale. The NMR data on isolated AT helix was collected by Sunayana Mitra. Many thanks also to my colleagues in Department of Biological Science for countless discussions, for advice on techniques and strategies. To Dr. Nathaniel Gilbert, Dr. Svetlana Pakhomova, Matthew Kobe, Dr. Lijian Xiao, Ambuj Kushwaha, Tyler Broussard, Yu Ge etc.

I also thank the financial supports from LONI institute and Department of Biological Science for funding my graduate education, and Biograds, McDaniel Award and National Biomedical Computation Resource for my travel to participate in conferences or workshops.

Finally, a special thanks to you for reading this dissertation about my tiny contributions the advancement of science. Comments are always welcome!
TABLE OF CONTENTS

DEDICATION ........................................................................................................ ii

ACKNOWLEDGEMENTS ...................................................................................... iii

LIST OF TABLES .................................................................................................... vii

LIST OF FIGURES .................................................................................................. viii

ABSTRACT .............................................................................................................. xi

CHAPTER 1. RIBOSWITCHES—EMERGING CIS-REGULATORY ELEMENTS ..........1
  1.1 Diverse Roles of RNA in the Central Dogma ........................................ 2
  1.2 Regulation of Gene Expression by Riboswitches .................................... 4
  1.3 Influence of Sequencing Technology and Computational Techniques on ncRNAs .... 7
  1.4 Riboswitches as Drug Targets for Novel Antibiotics ............................. 8
  1.5 Riboswitches as Portable Devices for Synthetic Biology ...................... 9
  1.6 The SAM Riboswitch Family ................................................................. 10
  1.7 The Scope of This Work ....................................................................... 14

CHAPTER 2. A MECHANISM FOR S-ADENOSYL METHIONINE ASSISTED FORMATION OF A RIBOSWITCH CONFORMATION: A SMALL MOLECULE WITH A STRONG ARM ......................................................... 17
  2.1 Updated Discussions ............................................................................ 18

CHAPTER 3. EXPLORING THE SECONDARY STRUCTURE FOLDING OF THE FULL LENGTH SAM-I RIBOSWITCHES USING IN SILICO CO-TRANSCRIPTION FOLDING SIMULATION (COTFS) ................................................ 20
  3.1 Introduction .......................................................................................... 21
  3.2 Materials and Methods ....................................................................... 22
  3.3 Results .................................................................................................. 26
  3.4 Discussions ........................................................................................ 51

CHAPTER 4. A ROLE OF SMALL MOLECULE IN STRAND SWITCHING FROM RNA MODELING AND MD SIMULATION .................................................. 59
  4.1 Introduction .......................................................................................... 60
  4.2 Structure Modeling Using MC-sym ...................................................... 63
  4.3 Results .................................................................................................. 67
  4.4 Discussions ........................................................................................ 84

CHAPTER 5. STRUCTURAL MODELING OF A SAM-I RIBOSWITCH IN THE “ON” STATE FROM NMR AND SAXS .............................................................. 89
  5.1 Introduction to Solution State Biophysical Methods ............................... 90
  5.2 Experimental Procedures .................................................................... 93
  5.3 Preliminary Results and Discussions .................................................... 96
CHAPTER 6 CONCLUSIONS AND FUTURE WORK ...........................................116
REFERENCES .................................................................................................121
APPENDIX A: GENOMIC CONTENT DOWNSTREAM OF SAM-I RIBOSWITCHES .....142
VITA .................................................................................................................144
LIST OF TABLES

Table 1.1 Information about six classes of SAM riboswitch, SAH riboswitch and SAM related riboswitch from Rfam\textsuperscript{55} ................................................................. 11

Table 3.1 Sequences of oligo templates and primers for PCR ...................................................... 23

Table 4.1 Summary of MD trajectories in this study ................................................................. 72

Table 5.1 R\textsubscript{g} and D\textsubscript{max} calculated based on the Guiner region of the 1D scattering profiles .... 115
LIST OF FIGURES

Figure 1.1 Active roles of RNAs in the central dogma. ..................................................4

Figure 1.2 Schematic of three regulatory mechanisms that riboswitches utilize to control gene expression. .................................................................6

Figure 1.3 Chemical structures of SAM and SAH. ............................................................12

Figure 1.4 Atomic details of SAM recognition by the SAM riboswitches. .........................13

Figure 1.5 Schematic of the regulatory mechanism of the SAM-I riboswitch. .................15

Figure 2.1 Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) calculation to decompose the total binding free energy into contributions from individual residues. ........17

Figure 2.2 Porcupine plot displayed the motions along the first eigenvector. .....................18

Figure 3.1 An example of RNA secondary structure. .........................................................21

Figure 3.2 Comparison of the secondary structure folding of two SAM-I riboswitch. ........27

Figure 3.3 Two possible secondary structural modes of the antiterminator (AT) helix in the “ON” form for T.tencongensis metF SAM-I riboswitch. ........................................29

Figure 3.4 Sequence alignment of SAM-I riboswitch sequences that are upstream of the metF gene in several thermophilic anaerobic bacterium. ........................................29

Figure 3.5 Sequence swapping test. ..................................................................................31

Figure 3.6 Heat maps of other B. subtilis SAM-I riboswitch sequences from reference 60. ....31

Figure 3.7 Plots of BPPs for base pairs potentially involved in strand switching at selected temperatures at a fixed transcript length allowing for P1 or AT base pair formation. ........33

Figure 3.8 Secondary structure representation of possible hybrid intermediate states. ..........35

Figure 3.9 Effects of G11 mutation examined by BPP calculation. ...................................36

Figure 3.10 Decoys in the junction region. .........................................................................38

Figure 3.11 Sequence survey for decoy interactions D1 and D2. ....................................40

Figure 3.12 BPP calculation with constraints mimic SAM binding. ..................................41

Figure 3.13 Gel image from in-line probing experiment. ...............................................44
Figure 3.14 Gel image from in-line probing experiment. ..............................................45

Figure 3.15 Native gel mobility of the \textit{T.tengcongensis metF} SAM-I riboswitch and mutants at different length of transcript. .................................................................46

Figure 3.16 Effects of transcript length on SAM binding affinity examined by equilibrium dialysis. ..............................................................................................................48

Figure 3.17 Steady state fluorescence experiment on \textit{T.tengcongensis} SAM-I riboswitch RNA constructs. .................................................................49

Figure 3.18 UV melting experiment of isolated P0 helix. ........................................51

Figure 3.19 Schematic illustration of 3 conformational decision points during the synthesis of the \textit{T. tengcongensis metF} SAM-I riboswitch. .....................................................53

Figure 4.1 Schematic of observations from SAM binding experiment. .........................62

Figure 4.2 Secondary structure representation of the hybrid construct--6P1_11AT. ........65

Figure 4.3 MC-sym sampling benchmark. ....................................................................69

Figure 4.4 Structural analysis of MC-sym models. ........................................................70

Figure 4.5 Structure analysis of MC-sym models. ..........................................................71

Figure 4.6 Time evolution of RMSD for individual base pairs in the P1 helix and the AT helix. ..................................................................................................................75

Figure 4.7 Monitor of hydrogen bonds in Watson-Crick base pairs. .........................76

Figure 4.8 RNA-ligand interaction distance monitor. ..................................................77

Figure 4.9 Radius of gyration of different SAM-I riboswitch models during simulations. 77

Figure 4.10 Monitor of a local geometry in the simulation of model 55 with SAM. ..........78

Figure 4.11 PCA of the P1 helix from multiple starting structures. ...............................80

Figure 4.12 2D projection of individual trajectory onto the first two eigenvectors for PCA of the P1 helix. ......................................................................................................................81

Figure 4.13 Energy landscape visualization for simulations on model51. .....................84

Figure 4.14 Energy landscape visualization from PCA of the P1 helix. .........................85
Figure 5.1 UV melting of isolated AT_U2 helix. .................................98

Figure 5.2 UV melting of isolated AT_U4. .................................99

Figure 5.3 Imino-imino proton region in NOESY NMR spectrum of different isolated AT helices and assignment of some imino protons. .........................................................100

Figure 5.4 Comparison of the 1D imino proton region of different AT helix constructs. ........101

Figure 5.5 NMR temperature course experiment for isolated AT helix with wild type loop and isolated AT_U2. .................................................................101

Figure 5.6 Comparison of melting profiles for an "ON" conformer and the "OFF" conformer. .................................................................103

Figure 5.7 Melting profiles for GGU_AT_U2 RNA construct under different ionic condition. .................................................................103

Figure 5.8 NMR experiments on GGAA_AT_U2 RNA construct. .........................105

Figure 5.9 Comparison of 1D imino proton spectrum for isolated AT_U2, GGAA_AT_U2 and isolated P4 helix at 15 °C. .................................................................106

Figure 5.10 Scattering profile of GAGG_AT_U2 RNA construct under two different buffer conditions. .................................................................108

Figure 5.11 Coarse grain model of an “ON” state conformer—GAGG_AT_U2 from SAXS experiment. .................................................................109

Figure 5.12 Modeling atomic structure using SAXS data in the presence of Mg$^{2+}$. ..........110

Figure 5.13 Top 8 rank structure models from indirect fitting to coarse grain electron density. .................................................................111

Figure 5.14 NMR experiments on isolated P1 helix and WT AAT RNA. .................112

Figure 5.15 1D imino spectrum of different RNA constructs. .........................113

Figure 5.16 SAXS scattering profiles and gel filtration column profiles for WT sequence with different 3’ end truncation points. .................................................................115

Figure 6.1 Schematic of thermodynamics cycle model for the SAM-I riboswitch and an example of oligo devices used in DNA computing. ....120
ABSTRACT

The chemical and physical properties of RNAs create a diverse functional portfolio to influence the functional outcomes of genomes. One scheme is to recognize cognate small molecule metabolites and to adopt distinct conformations to adjust the gene expression. Segments of messenger RNAs (mRNAs) that adopt this scheme are called riboswitches. As potential targets for designing novel antibiotics and portable regulatory devices for synthetic biology, riboswitches have gained increasing attention. The key to understand the functionality encoded in a riboswitch sequence requires unveiling the mechanism of transmitting ligand recognition to gene expression.

In this work, both computational and experimental techniques are employed to investigate the link between cognate ligand binding and conformational rearrangement of the SAM-I riboswitch. This riboswitch modulates the biosynthetic pathways of methionine, cysteine, S-adenosylmethionine (SAM) and other sulfur containing metabolites at the transcriptional level. Molecular Dynamics (MD) simulation, with improved force field, extended time scale and empowered by advanced computer hardware, is used to explore the conformational dynamics in 3D space. A partition function approach is adopted to examine potential conformational heterogeneity within the functional decision windows of the SAM-I riboswitch during the synthesis of the transcript. A proposed framework combining RNA tertiary structure prediction with experimental observations and MD simulations appears to be plausible for modeling transient events during RNA folding. Finally, experimental techniques, such as chemical probing, equilibrium dialysis, UV melting, NMR spectroscopy and SAXS, are used to verify insights
gained from computational work or to generate structural information for further computational structure modeling.

This work has the following implications: 1) SAM plays an important role in anchoring the junction between helices one and two (J1/2), facilitating formation of the "OFF" state conformer, which may be synergized with formation of a nearby Mg$^{2+}$ binding site. 2) Alternative or "misfolded" conformations due to the interactions between J1/2 and decoy regions representing the "ON" state ensemble facilitate fine tuning of the SAM-I riboswitches. 3) Simulated strand switching within hybrid intermediate structures in the presence of SAM reveals atomic level details of SAM-induced stabilization of the transcriptional OFF state.
CHAPTER 1 RIBOSWITCHES—EMERGING C辐射-REGULATORY ELEMENTS
1.1 Diverse Roles of RNA in the Central Dogma

The discovery that RNA can catalyze chemical reactions without the help of protein\(^1,2\) started a paradigm shift. RNA is not simply a passive messenger. The high resolution structure of the ribosome\(^3,4\) has unveiled the fact that the catalytic core is solely composed of RNA. Therefore, the ribosome, one of the most important types of biological machinery, is also a ribozyme\(^5,6\). Moreover, there is an exploding realization of the importance of non-coding RNA (ncRNA)\(^7,8\). Back in 1961, Jacob and Monod predicted that “Structural genes encode proteins, and regulatory genes produce ncRNA.”\(^9\) The current trend of interest in ncRNA was initiated by the surprising conclusion of the human genome project that the number of protein coding genes is 20,000-25,000\(^10\), which is only \(~2\%\) of the total transcripts\(^11\). The following is a list of some ncRNA families that are involved in different activities in the central dogma:

- **Transfer RNA (tRNA), ribosomal RNA (rRNA):** tRNA and rRNA are the two most well known RNA families. tRNAs act as an adaptor between mRNA and amino acid in protein synthesis. rRNAs are the structural and the catalytic component of ribosome\(^3\).

- **Ribozyme:** Ribozymes are enzymes in RNA form, such as group I introns, group II introns etc.

- **Spliceosomal RNA:** Spliceosomal RNA is the RNA component of the large ribonucleoprotein complex—spliceosome, including five RNA species (snRNAs U1, U2, U4, U5 and U6)\(^12\).

- **Long noncoding RNA (lncRNA):** lncRNAs are ncRNAs that are greater than 200 nucleotides in length\(^13\). LncRNAs have been shown to be involved in transcriptional regulation, epigenetic gene regulation and diseases\(^14\).
• **Small interfering RNA (siRNA):** siRNAs are 20-25 residue double strand RNAs (dsRNAs) that cause interference with gene expression\(^{15-17}\). Since the early days when siRNAs were discovered, siRNAs have been widely used to manipulate gene expression in biomedical research \(^{16}\), and they have also become a promising tool for gene-specific therapeutics \(^{18}\).

• **Micro RNA (miRNA):** miRNAs are also double strand RNAs with ~23 nt at length, and play important roles in gene regulation in animals and plants \(^{19}\). miRNA is different from siRNA in terms of its biogenesis \(^{20}\). The precursor of miRNA is ssRNA in the form of a hairpin secondary structure. The other difference is that typically miRNA is not completely complementary to its target \(^{20}\).

• **Small nucleolar RNA (snoRNA):** snoRNAs mediate the site-specific modification of other RNAs, mostly rRNA \(^{21}\). There are two classes of snoRNAs—C/D box and H/ACA box, which are involved in 2’-O-ribose methylation and pseudouridylation modifications respectively.

• **Piwi-interacting RNA (piRNA):** piRNA is a set of 24-30mer small RNAs that associate with P-element Induced WImpy testes (PIWI) protein \(^{22}\). piRNAs have been shown to be involved in protecting genomes from transposable elements \(^{23}\).

• **Transfer-messenger RNA (tmRNA):** tmRNA, also named SsrA RNA or 10Sa RNA, functions both as a tRNA and as an mRNA. tmRNA play a critical role in the quality control of protein biosynthesis and in recycling ribosomes that are stalled or interrupted by defective mRNAs in bacteria \(^{24}\).

• **Riboswitch:** Riboswitches are cis-regulatory elements that locate at the 5’ untranslated region (UTR) of the mRNA that mediate gene expression by sensing the level of
metabolites in bacteria. So far, there is only one riboswitch—thiamine pyrophosphate (TPP) riboswitch found in eukaryotes.

Therefore, a myriad of functionally diverse RNAs play critical roles in a revised central dogma (Fig. 1). Instead of sitting in the middle as intermediate information storage, RNAs are now understood to be play active roles in gene regulation.

Figure 1.1 Active roles of RNAs in the central dogma.

1.2 Regulation of Gene Expression by Riboswitches

Riboswitches have been shown to be key players in the gene regulation of bacteria. A riboswitch is an autonomous cis-regulatory element that typically locates at the 5’ untranslated region (UTR) in bacteria. The physiological signals that trigger the regulation via riboswitches...
can be small molecule metabolites\textsuperscript{35-40} and ions\textsuperscript{41}, temperature\textsuperscript{42} and pH\textsuperscript{43}. A riboswitch is typically comprised of an aptamer domain and an expression platform. The aptamer domain includes the part of the RNA sequence that a small molecule binds to. The expression platform is the region where a structural arrangement occurs when a riboswitch encounters its cognate ligand. Thus, it facilitates the transmission of detected physiological signals into control of gene expression. Figure 2 demonstrates the three most common mechanisms that riboswitches adopt to exert their functions. Regulation at the transcriptional level and the translational level are achieved by affecting the formation of a rho-independent terminator structure (a hairpin RNA with a stretch of Us at the 3’ end) and the sequestration of the Shine-Dalgarno (SD) sequence (a ribosome binding site to initiate translation) respectively. In some cases, a riboswitch can also be a ribozyme—a self-cleavage occurs upon the binding of the cognate ligand, and thus the susceptibility to RNases is changed to reduce the half-life of the mRNA.

Riboswitches manifest their capacity as precise regulatory elements. This is because of their high specificity for their cognate physiological signals—the capacity that enables riboswitches to discriminate between closely related metabolites so as to reach the correct output from the environmental stimulus. Additionally, it has been shown that slight sequence variability allows riboswitches within the same family to finely tune the level of genes. The products of these genes are involved in different stages of biosynthetic pathways\textsuperscript{44}. There are also cases of multiple riboswitches from the same family\textsuperscript{45} or from different families\textsuperscript{46} are tandemly arranged to modulate gene expression in a composite mode.

Most known riboswitches currently have only been identified in bacteria. However, there is an exception. TPP riboswitches have been characterized in all three domains of life\textsuperscript{26,28,36,47-49}. In eukaryotes, TPP riboswitches modulate gene expression by altering the splicing site in
response to TPP. The proof of TPP riboswitch function in eukaryotes raises the possibility that more instances of riboswitch-like gene regulation are likely to exist in eukaryotic organisms. The fact that the other bacterial riboswitch families have not been identified in any eukaryote to date may be due to the evolution of novel variations of this ancient gene control mechanism \(^50\). The evolution of new mechanisms in eukaryotes will then require different strategies for identifying this type of genetic element in the eukaryote. Recently, a similar genetic regulatory mechanism that utilizes the conformational flexibility of RNA has been described in human \(^51,52\). Therefore,
riboswitch-like regulatory elements—functioning by switching between alternative RNA conformations—may represent a much more common genetic regulatory mechanism than is currently suspected.

1.3 Influence of Sequencing Technology and Computational Techniques on Studying ncRNAs

The advancement of DNA sequencing technology (massively parallel sequencing) and computational hardware and algorithms (parallel computing) has greatly accelerated the growth of genomic and metagenomic sequence data. The results and related information have been organized into primary databases, for example, the Genomes OnLine Database (GOLD). The growth of the primary database also enlarges the volume of the secondary database, for example, Rfam. Rfam is a database to deposit RNA families with curated RNA structural alignments and consensus structures. One of the obvious benefits for researchers is that instead of working in the “wet” lab to figure out the identity of targeting RNAs, they can start their research by pulling out sequencing information from these online databases. The availability of these resources has facilitated the exploration and discovery of well structured ncRNAs from the level of DNA sequences.

Secondly, deep coverage and nucleotide resolution RNA-seq sequencing techniques in combination with computational methods have rapidly enlarged the ncRNA families. These techniques have demonstrated the ability to explore the structural features of RNA in genome-wide measurement. Lu et al. have developed a method integrating multiple computational methods, high-throughput sequencing and array data to separate UTRs from protein coding regions so as to discover novel ncRNAs. Additionally, several recent studies have designed strategies to couple different RNA structure probing experiments, such as RNase cleavage
mapping and SHAPE probing\textsuperscript{64}, with high-throughput sequence techniques to yield genome-wide RNA structural information from heterogeneous RNA samples. Altogether, these studies present the promise of expanding our understanding of ncRNAs using high-throughput sequencing and advanced computational methods.

1.4 Riboswitches as Drug Targets for Novel Antibiotics

The emergence of multiple antibiotic resistance still calls for the development of novel antibiotics\textsuperscript{65}. According to CDC, there are 15,249 mortalities caused by Methicillin-Resistant \textit{Staphylococcus aureus} infection in the United States in 2008\textsuperscript{66}. There are three major mechanisms that bacteria utilize to evade a drug class\textsuperscript{67}. One of the mechanisms is that bacteria use a multiple resistance efflux system to prevent the drug from reaching the intracellular target. Additionally, bacteria express enzymes that can modify antibacterial drugs so as to achieve resistance. A third strategy of resistance is that bacteria can modify the target while maintaining its physiological activity to deceive the antibacterial agents. The acquisition of resistance is attributed to chromosomal mutations or horizontal gene transfer\textsuperscript{68}.

One of the strategies to discover novel antibiotics and to circumvent the resistance problem is to find new bacterial drug targets\textsuperscript{67}. Riboswitches hold promise for their potentials as antibacterial drug targets\textsuperscript{69,70}. First, the aptamer domain of riboswitches can fold into well-defined tertiary structures\textsuperscript{71-80} with high specificity\textsuperscript{35,81-84} for metabolites that have drug-like properties. This also facilitates some structure-based drug design\textsuperscript{85} and docking\textsuperscript{86} studies on targeting riboswitches. Secondly, riboswitches are involved in modulating the expression of genes in important metabolic pathways. It is believed that interruption of normal gene production by targeting riboswitches could give rise to deficits of essential metabolites or accumulations of toxic intermediate metabolites\textsuperscript{69}. Therefore, this strategy can inhibit the growth
of bacteria or even kill the bacteria. Thirdly, antibiotics that target riboswitches may avoid the quick gain of resistance in bacteria. Mutations that affect the binding of the small molecule drug also cause the functional loss of riboswitches. The idea that the regulation function is lost as a result of drug insensitive mutations has been exemplified in several recent studies.

In the past several years, progress and lessons in the field have solidified the support for targeting riboswitches and paved the way for future novel antibiotic discovery. The first study to provide proof of principle identified several L-lysine analogues that bind to the lysine riboswitch and inhibit the growth of B. subtilis. However, a study later on has demonstrated that the lysine riboswitch is not the primary target for these amino acid antibiotics, suggesting that the acquisition of resistance is due to mutations in protein targets. The trend of targeting riboswitches also led to the discovery that a natural antibacterial compound—roseoflavin actually exerts its antimicrobial function through repressing genes in riboswitch-mediated pathways. With well characterized three dimensional structures and extensive biochemical and biophysical studies, the discovery of new ligands for purine riboswitches has been accelerated. Several recent studies on this specific target yielded several promising compounds. Moreover, one study has for the first time demonstrated that an antibiotic targeting a riboswitch is quite effective in treating infections with bacterial pathogens in mammalian models.

1.5 Riboswitches as Portable Devices for Synthetic Biology

Synthetic biology is an emerging field that uses biological components to engineer artificial genetic circuits so as to acquire novel functions for biological systems. The idea of utilizing small molecules to tune the gene expression through their interactions with RNA has come of age. Furthermore, the fact that riboswitches can serve as precise genetic control
elements has inspired the use of riboswitches to construct synthetic biological systems. There are several advantages of using riboswitches to construct synthetic biological systems. First, riboswitch mediated regulatory is simplified compared to other signal cascades using proteins or hybrid protein/RNA by eliminating the step of protein translation. This feature will allow the construction of a large scalable gene network. Additionally, cell-permeable small molecules can be introduced externally as the stimulus signals. A recent study has utilized this feature of riboswitches to construct a novel biological system to degrade contaminants in water. Furthermore, nature has displayed the possibility of arranging tandem riboswitches to achieve complicated signal gating control. The construction of genetic devices to process high-order cellular information can also be achieved using different strategies to assemble riboswitches with ribozymes. It is also possible to finely tune the regulation with minor sequence variation. However, it is important to note that currently natural and synthetic riboswitches only selectively recognize a limited number of small molecule ligands. Several strategies have been proposed to enlarge the arsenal of riboswitches to circumvent this challenge by designing high-throughput screening for synthetic riboswitches starting with RNA aptamers from SELEX and by orthogonal evolution from known natural riboswitches.

1.6 The SAM Riboswitch Family

S-adenosyl methionine (SAM) (Figure 1.3) is a very special metabolite. SAM is the second most popular enzyme cofactor only after ATP. SAM is the major source of methyl group in the methylation reaction. SAM also acts as an “organ donor” in a wide spectrum of chemical reactions. It is also noticeable that S-adenosyl homocysteine (SAH), the resulting byproduct when SAM donates the methyl group, is toxic to cells. The special significance of SAM hints at the requirement for the stringency of regulation. Indeed, the SAM riboswitch is
Table 1.1 Information about six classes of SAM riboswitch, SAH riboswitch and SAM related riboswitch from Rfam

<table>
<thead>
<tr>
<th>ID</th>
<th>Rfam Accession</th>
<th>Type</th>
<th>Seed†</th>
<th>Full‡</th>
<th>Average Length</th>
<th>Sequence Identity (%)</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>SAM</td>
<td>RF00162</td>
<td>Cis-reg; riboswitch</td>
<td>228</td>
<td>2,828</td>
<td>109.10</td>
<td>69.00</td>
<td>SAM riboswitch (S-box leader)</td>
</tr>
<tr>
<td>SAM alpha</td>
<td>RF00521</td>
<td>Cis-reg; riboswitch</td>
<td>40</td>
<td>227</td>
<td>79.00</td>
<td>71.00</td>
<td>SAM-II riboswitch (alpha-proteobacteria)</td>
</tr>
<tr>
<td>SAM-II long loops</td>
<td>RF01726</td>
<td>Cis-reg; riboswitch</td>
<td>127</td>
<td>236</td>
<td>60.70</td>
<td>77.00</td>
<td>SAM-II long loops</td>
</tr>
<tr>
<td>SMK_box</td>
<td>RF01767</td>
<td>Cis-reg; riboswitch</td>
<td>25</td>
<td>72</td>
<td>86.50</td>
<td>71.00</td>
<td>SAM-III riboswitch (SMK box)</td>
</tr>
<tr>
<td>SAM-IV</td>
<td>RF00634</td>
<td>Cis-reg; riboswitch</td>
<td>40</td>
<td>267</td>
<td>115.80</td>
<td>73.00</td>
<td>SAM-IV riboswitch</td>
</tr>
<tr>
<td>SAM-V</td>
<td>RF01826</td>
<td>Cis-reg; riboswitch</td>
<td>6</td>
<td>722</td>
<td>65.80</td>
<td>80.00</td>
<td>SAM-V riboswitch</td>
</tr>
<tr>
<td>SAM-I/IV variant</td>
<td>RF01725</td>
<td>Cis-reg; riboswitch</td>
<td>439</td>
<td>628</td>
<td>99.00</td>
<td>58.00</td>
<td>SAM-I/V variant riboswitch</td>
</tr>
<tr>
<td>SAH</td>
<td>RF01057</td>
<td>Cis-reg; riboswitch</td>
<td>52</td>
<td>225</td>
<td>84.70</td>
<td>63.00</td>
<td>SAH riboswitch</td>
</tr>
<tr>
<td>SAM/SAH</td>
<td>RF01727</td>
<td>Cis-reg; riboswitch</td>
<td>63</td>
<td>76</td>
<td>48.70</td>
<td>73.00</td>
<td>SAM/SAH riboswitch</td>
</tr>
<tr>
<td>SAM-Chlorobi</td>
<td>RF01724</td>
<td>Cis-reg; riboswitch</td>
<td>9</td>
<td>12</td>
<td>89.00</td>
<td>74.00</td>
<td>SAM-Chlorobi tandem riboswitch</td>
</tr>
</tbody>
</table>

*Cis-reg indicates that the riboswitch has been characterized only through bioinformatics analysis; riboswitch means that experiment has been performed to verify the regulatory function.
† Seed alignment is a manually curated alignment that contains representative members of the ncRNA family and is annotated with structural information.
‡ Full alignment is a result of search using the covariance model created using seed alignment against the sequence databases.
currently the largest riboswitch family among known classes of riboswitches, which includes six distinct classes of SAM-binding riboswitches $^{119-124}$—SAM-I, II, III, IV, V and I/IV variant (Table 1). The SAM riboswitch family is widely distributed in different bacterial phyla $^{32}$, which may be due to the importance of sensing this ligand in cells. The SAM-I riboswitch is the most prevalent one among different classes of SAM riboswitches $^{32}$. It has been shown that these SAM riboswitches utilize different regulatory mechanisms to exert their functions $^{32}$, for example, the SAM-I riboswitch mainly functions at the transcriptional level $^{125}$, while the SAM-III riboswitch functions at the translational level $^{121}$. There are also some cases in which the SAM riboswitch is arranged consecutively with another riboswitch, for example, the coenzyme B12 sensing riboswitch $^{46}$. Additionally, there is a class of SAH riboswitch involved in modulating genes involved in SAH recycling pathways to prevent the accumulation of this toxic metabolite $^{118}$. The SAM riboswitch family and SAH riboswitch have displayed the capability of discrimination between closely related metabolites by riboswitches—SAM riboswitch has strong discrimination in favor of SAM $^{39,81,82}$, while SAH riboswitch binds to SAH more tightly than SAM $^{118}$. The SAM riboswitch family also demonstrates the structural versatility of RNA as it evolves to form selective binding pockets for the specific metabolite.

![Figure 1.3 Chemical structures of SAM and SAH.](image-url)
Figure 1.4 Atomic details of SAM recognition by the SAM riboswitches. Tertiary structure of SAM-I riboswitch (PDB: 2GIS) (a), SAM-II riboswitch (PDB: 2QWY) (b) and SAM-III riboswitch (PDB: 3E5C) (c) in cartoon representation with two different views (Left) and (Middle) by rotation along the y-axis. (Right) Local view of the electrostatic interactions between the positively charged sulfur atom of SAM and the negatively charged oxygen atoms of RNA, and their positions in the whole structure are highlighted in red rectangle box (Middle). The number in the (Right) panel is the distance between two atoms with the unit—Angstrom. The notation P stands for the helical region, J stands for the junction region and L is short for loop.
Several crystal structures for the aptamer domain of SAM riboswitches (I, II and III) and a SAH riboswitch have been solved \cite{75,79,82,87,126,127}. These structures have revealed the details of molecular recognition at the atomic-level (Figure 1.4). The strong discrimination between SAM and SAH by the SAM-I riboswitch is achieved by the electrostatic interactions between the positively charged sulfur atom of SAM and the carbonyl oxygen atoms (O2) of two U residues in the RNA \cite{79}. Similarly, the SAM-II riboswitch utilizes the carbonyl oxygen atoms (O4) of two U residues to specifically select the positive charge on sulfur \cite{75}. Again likewise, the SAM-III riboswitch employs a carbonyl oxygen atom (O4) of a U residue and a hydroxyl oxygen atom (O2') of a G residue to attract the sulfur atom of SAM \cite{126}. In the case of the SAH riboswitch, formation of a binding pocket that sterically hinders the additional methyl group of SAM is the basis to favor SAH over SAM \cite{118}.

This work focuses on the studies of the SAM-I riboswitch. The SAM-I riboswitch modulates expression of genes involved in the biosynthesis of methionine, cysteine, SAM and other sulfur-containing metabolites in response to the level of SAM in vivo \cite{39,128,129}. Genetic and biochemical studies propose the following model for the regulatory mechanism of SAM-I riboswitch (Figure 1.5). In the absence of SAM, the SAM-I riboswitch forms the anti-terminator (AT), which prevents formation of the terminator (T) that is required for rho-independent transcription attenuation. Binding by SAM triggers formation of the anti-anti-terminator (AAT) (henceforth referred to as the aptamer) that occupies the 5’ strand of the AT helix. Under these conditions, the formation of the T facilitates transcription termination, turning off the expression of a set of downstream genes that are involved in sulfur-related metabolic pathways.

1.7 The Scope of This Work

The functional mechanism of riboswitches allows changes in metabolite concentration to
Figure 1.5 Schematic of the regulatory mechanism of the SAM-I riboswitch.

orchestrate the level of a set of enzymes in a related biosynthetic pathway. The critical step of this functional mechanism is the transmission of ligand binding information to the expression platform mainly through conformational arrangement. Understanding the basis of this information flow can facilitate the usage of riboswitches as portable regulatory elements for synthetic biology. Thus, this work requires studying the riboswitch in the full-length sequence context and extracting dynamic information. Both computational and experimental techniques have been used toward these goals. In Chapter 2, Molecular Dynamics (MD) simulations are employed to present a dynamic view of the aptamer domain of the SAM-I riboswitch in the free state and the bound state. From this work, we propose hypotheses regarding the role of specific contacts between ligand and RNA at the atomic level, in facilitating this signal transfer. In Chapter 3, we extend the sequence into the expression platform and utilize secondary structure prediction with a partition function approach to identify conformers associated with different functional states. Meanwhile, experimental results from chemical probing, equilibrium dialysis and fluorescence experiment in Chapter 3 provide indirect evidence for the formation of alternative conformation. Further sequence analysis has verified some transient interactions and
provides insight the functional significance of these alternate forms. We propose roles for these minor conformers in modulating riboswitch conformation and related function. In Chapter 4, rational modeling with experimental observation and trajectories at unprecedentedly long time-scale for RNA of this size has provided atomic details of a plausible pathway for the information transmission. This work enables us to speculate as to how SAM interactions with the alternate/minor states could play a role in mediating SAM-induced riboswitch folding. In Chapter 5, UV melting, NMR spectroscopy and SAXS experimental techniques lay out a foundation to characterize the dynamics of the full-length SAM-I riboswitch.
CHAPTER 2 A MECHANISM FOR S-ADENOSYL METHIONINE ASSISTED FORMATION OF A RIBOSWITCH CONFORMATION: A SMALL MOLECULE WITH A STRONG ARM\textsuperscript{130§}

\textsuperscript{§} Author’s Journal Article in \textit{Nucleic Acids Research}
2.1 Updated Discussions

MM-GBSA has been used to quantitatively evaluate the interactions between SAM and individual residues\textsuperscript{131}. Figure 2.1 shows the decomposition of the calculated total binding free energy into the contributions from individual residues. The results show that G58 in the P3 helix and G11 in the J1/2 junction are the two residues that have the largest contributions to the SAM binding, and C47 in the P3 helix also makes significant contributions to the binding. These results are consistent with loss of SAM binding in single residue mutants of corresponding residues (G11, G79 and C45) in the context of \textit{yitJ} SAM-I riboswitch\textsuperscript{87}. Even the mutant C45U (\textit{yitJ}) that would replace the G79-C45 base pair with a GU base pair eliminates the SAM binding. However, the mutant A111G (\textit{yitJ}), which allow a GU base pair to replace the A111-U7 base pair in the P1 helix, only results in decreased binding affinity. Altogether, these results suggest that the P3 helix and the J1/2 junction are crucial for organizing a proper scaffold for the SAM binding.

![Figure 2.1](image.png)

**Figure 2.1** Molecular Mechanics-Generalized Born Surface Area (MM-GBSA) calculation to decompose the total binding free energy into contributions from individual residues.

Figure 2.2 shows the porcupine plot describing different motions observed in the aptamer simulations with and without SAM. The motion observed in the absence of SAM, termed the
“chopstick” motion, is consistent with structural modeling based on SAXS data demonstrating a “scissor” like motion between the P1 and the P3 helices\textsuperscript{132}. The dynamics of the P4 helix is coupled with the P1 helix in the absence of SAM as indicated by directions of arrows shown in Figure 2.2b, while this motion is not observed in the presence of SAM which displays an opposite trend. The dependence of the P4 helix dynamics on SAM is also in line with a recent single molecule study\textsuperscript{133}.

![Porcupine plot](image)

**Figure 2.2** Porcupine plot displayed the motions along the first eigenvector. The first eigenvector is visualized as arrows mapping onto the backbone atom O5’ of the average structure. (a) MD simulation with SAM, (b) MD simulation without SAM. Projection of the trajectory onto the eigenvector shows the trend of the motion as indicated with the arrows. The relative length of the arrows shows the relative magnitude of the fluctuations extracted from the trajectory for each mapped atom. This plot is used to visualize the large-scale motion, a large number of atoms moving in concert, observed from the simulations. In the presence of SAM, the large-scale motion occurs in the P4 helix. While in the absence of SAM, P1 and P3 helices are moving in the opposite direction—the chopstick motion as described in reference\textsuperscript{130}.
CHAPTER 3 EXPLORING THE SECONDARY STRUCTURE FOLDING OF THE FULL LENGTH SAM-I RIBOSWITCHES USING \textit{IN SILICO} CO-TRANSCRIPTION FOLDING SIMULATION (COTFS)
3.1 Introduction

The folding of RNA follows a hierarchy route—more stable secondary structure forms first and is followed by the tertiary folding. The hierarchy folding model suggests that secondary structure folding is independent of tertiary folding. RNA secondary structure is composed of 2D structural motifs—hydrogen bonded double strand helices, terminal mismatch, bulge, internal loop, hairpin loop and multi-branch loop (Figure 3.1). The thermodynamic parameters for these structural elements can be determined through thermodynamic measurements and improved by machine learning techniques. The free energy change upon RNA folding is then estimated with a nearest neighbor model by summing the free energy increments from each structural element. Based on this nearest neighbor model, a dynamic programming algorithm is used to search for the minimal free energy (MFE) structure or suboptimal structures. This method, so called free energy minimization, can correctly predict ~70% of secondary structures for RNA sequences with fewer than 7000 nucleotides.

Figure 3.1 An example of RNA secondary structure.
Free energy minimization methods only yield a single MFE structure or several sub-optimal structures, which are less informative in some cases. Alternatively, a partition function can be used to explore the ensemble of secondary structures for an RNA system. The partition function, $Q$, describes the statistical properties of a system in thermodynamic equilibrium. In the case of secondary structure folding, $Q$ is defined as the sum of the equilibrium constants for all possible secondary structures of a RNA sequence $^{144}$. From $Q$, a metric—base pair probability (BPP) can be calculated to provide more complete and reliable structural information. For RNA secondary structure prediction, McCaskill’s approach $^{145}$ for partition function calculation can be used to extract the BPP. BPP has been used in a study to identify region that can form well-defined secondary structure in the whole genome of an HIV virus RNA $^{146}$. Recently, BPP has also been used to evaluate the base pairing in the MFE structure and to assemble structures, termed as maximum expected accuracy structure, with only base pairs above a certain threshold of predicted BPP $^{144}$.

The objective of this study is to explore the secondary structure folding of riboswitches using sequence segments incorporating sections of the expression platform. In this way, we aim to decode the information flow from aptamer domain to the expression platform. The secondary structure folding of two SAM-I riboswitches upstream to the same gene but from different organisms are compared. BPP is used to estimate the population of RNA conformers for two alternative states—“ON” and “OFF”. The temperature dependence of BPP is also examined.

3.2 Materials and Methods

3.2.1 Sequence Source and RNA Preparation

The DNA templates for *T. tengcongenesis metF* SAM-I riboswitches were amplified by
overlapping PCR using oligos purchased from Integrated DNA Technologies, Inc.. The sequences of oligo templates and primers are listed in Table 3.1. RNA was transcribed from the PCR product by *in vitro* transcription using the following recipe: 100 µL PCR product, 50 µL 40 mM NTPs mix, 50 µL 400mg/mL PEG8000, 12.5 µL 20x transcription buffer (800 mM Tris-HCl, pH 8.1, 6 mg/mL spermidine, 0.2% Triton x-100, 15.6 mg/mL DTT), 11.5 µL 1M MgCl₂, 2 µL 4 mg/mL in house T7 RNA polymerase, add ddH₂O to 250 µL. Transcription and subsequent purification of the RNA was as described. The RNA sample was exchanged into 10 mM K Phosphate (pH 6.0), 10 mM KCl buffer and concentrated using Amicron. The RNA sample was stored at -20 °C.

**Table 3.1 Sequences of oligo templates and primers for PCR**

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Top Strand Template</strong></td>
<td>5- GCC AGT GAATTC TTA AAA TCT CTT ATC AAG AGA GGT GGA GGG ACT GGC CCG ATG AAA CCC GGC AAC CAG CCT TAG GGC ATG GTG CCA ATT CCT GCA -3</td>
</tr>
<tr>
<td><strong>Bottom Strand Template</strong></td>
<td>5- AAA GCC ATG GCC GGC CGC TAA AAG AAG AGA AGA CTA CAA GAATCT CTC ATC TTT CAG CGA AAC CGCTGC AGG AATTGG CAC CAT GCC CTA AGG CTG -3</td>
</tr>
<tr>
<td><strong>WT Forward Primer</strong></td>
<td>5- TAATAC GACTCA CTATAG AAT CTC TTATCA AGA GAG GTG GAG GG -3</td>
</tr>
<tr>
<td><strong>dP0 mutant Forward Primer</strong></td>
<td>5- TAATAC GACTCA CTATAG AAT CTC TTATCC AGA GAG GTG GAG -3</td>
</tr>
<tr>
<td><strong>eP0 mutant Forward Primer</strong></td>
<td>5- TAATAC GACTCA CTATAG AAT CTC TTATCC AGA GAG GTG GAG -3</td>
</tr>
<tr>
<td><strong>Aptamer (AAT) Reverse Primer</strong></td>
<td>5- mAmAT CTC TCA TCT TTC AAC GAA ACC GCT GCA GG -3</td>
</tr>
<tr>
<td><strong>AT1 Reverse Primer</strong></td>
<td>5- mAmAG AAG AGA CTA CAA GAATCT CTC ATC TTT CAG CG -3</td>
</tr>
<tr>
<td><strong>AT2 Reverse Primer</strong></td>
<td>5- mCmUT CGCTAA AAG AGA CTA CAA GAATCT CTC ATC TTT CAG -3</td>
</tr>
</tbody>
</table>
3.2.2 Base Pair Probability Calculation

BPP was calculated using McCaskill’s algorithm implemented in Vienna RNA Package 1.8.5 with or without constraints. BPP for specific base pairs were extracted from the dot plot files using in-house scripts. The net base pair probability for a single nucleotide was obtained by calculating the sum of base pair probabilities that the nucleotide was involved in (BPSUM). The number of possible base pairing partners (BPNUM) and the maximum BPP (BPMAX) among these base pairs were also extracted. These should be understood as predicted quantities throughout the manuscript. The scripts are available in the following link:

http://dl.dropbox.com/u/24028147/CTFS_v1.0.tar.gz

3.2.3 Equilibrium Dialysis

The RNA sample was heated at 95 °C for 3 minutes and snap cooled on ice for 5 minutes. Then the RNA was exchanged into 50 mM Tris-HCl (pH 7.5), 25 mM NaMES, 25 mM NaOAc, 100 mM KCl, 2 mM MgCl₂ to achieve a specified final concentration. 25 µL of the RNA sample was added to the chamber B of an equilibrium dialysis chamber—DispoDialyzer (Harvard Apparatus), and then the specified concentration of [³H] SAM was added to the chamber A of the dialyzer. The dialyzers were agitated on a nutator for ~8 hours at different temperatures (4 °C & RT/~22°C). Aliquots in the control (no RNA in chamber B) were collected for scintillation counting to confirm the equilibrium had been reached. The ratio of c.p.m of chamber B relative to chamber A, averaged over 3 readings, is reported as a measure of the relative binding affinity.

3.2.4 In-line Probing

In-line probing assays were performed following instructions in ref. The procedures
are briefly described here. The RNA was 5’ end labeled with $\gamma$-$^{32}$P ATP and gel purified. The in-line probing assays were performed under the following buffer condition: 50 mM Tris-HCl, 100 mM KCl (pH 8.3). The reactions were carried out with different Mg$^{2+}$ concentration (0, 2 mM, 10 mM, 40 mM), in the absence or in the presence of 1, 10 and 100 $\mu$M SAM. The results are quantified using the software—SAFA $^{154}$.

3.2.5 Steady State Fluorescence Experiment

The single strand RNA of 5’P1 helix with PyrroloC was obtained from the Keck Oligo Synthesis Lab in Yale University. The RNA is 2’ deprotected and desalted following the protocol from Glen Research (http://www.glenres.com). The steady-state fluorescence measurement is performed on a JASCO-FP6200 spectrofluorometer. The fluorescence intensity is recorded by exciting PyrroloC at 337 nm (5 nm band width) and monitoring the emission at 450 nm (10 nm band width) $^{155}$. The buffer condition used in this fluorescence experiment is 50 mM Tris, 25 mM MES-NaOH, 25 mM acetic acid-NaOH at pH 7.5, 100 mM NaCl. The experiment is performed at room temperature (22 °C).

3.2.6 Native Gel Assay

The RNA stock was diluted to 50 $\mu$M using 10 mM K Phosphate (pH 6.0), 10 mM KCl. The RNA sample was heated at 95 °C for 3 minutes and snap cooled on ice for 5 minutes. The RNA was folded in 1x TB (pH 8.3), 200 mM KCl, 10 mM MgCl$_2$, 20% glycerol with a final concentration of 5 $\mu$M and incubated at 37 °C for 1 hour with or without 100 $\mu$M SAM. The samples were electrophoresed using 10% native PAGE in 1x TB with 10 mM MgCl$_2$ at 4 °C at 200 V for 16 hours. This type of native gel assay has been used to select the conformation that has response to SAM in reference$^{156}$. 
3.3 Results

3.3.1 Different Secondary Structure Folding Behaviors of Two SAM-I Riboswitches

Atomic resolution X-ray structures have been reported for SAM-I riboswitches from one thermophilic \(^{157,158}\) and one mesophilic \(^{159}\) organism. For these systems we refer to the ligand-bound state, which correlates with reduced biosynthesis, as the “OFF” state. The unbound state, which permits increased biosynthesis, is called the “ON” state. For the two sequences rather different secondary structure models for the “ON” state are presented in the literature \(^{156,157,160}\). In particular, the two models differ in the length and characteristics of the so-called “antiterminator” (AT) helix that competes against the formation of a rho-independent terminator hairpin (Figure 3.2). These differing “ON” state structures might be linked to differing functional behaviors for SAM-I riboswitches upstream of genes involved in different components of sulfur or methylation-related metabolism. We chose to explore what differences in distribution of conformer populations would be predicted based on BPP calculations.

We input a series of RNA sequences of increasing transcript lengths for secondary structure prediction, incrementing the 3’ truncation point one nucleotide at a time. In this manner, we aim to predict folding intermediates as they may evolve during the synthesis of the riboswitch-containing region \(^{161}\). We call this type of calculation “co-transcriptional folding simulation” (CTFS). Then we monitored the BPP of the P1 helix as a function of temperature for each transcript length. The SAM-I riboswitch seems particularly suited to this analysis because a wealth of experimental binding and folding data is available \(^{156,158-160,162-166}\) particularly for sequence variants. The structural data and folding predictions suggest that relatively few non-canonical base pairs, (aside from tandem GA pairs in the well-characterized “kink-turn” domain \(^{167}\) ) are present. One weakness of secondary structure predictions, the lack of parameters
Figure 3.2 Comparison of the secondary structure folding of two SAM-I riboswitch. (a) & (b) Sequences of *B.subtilis yitJ* and *T.tengcongenesis metF* SAM-I riboswitches in the secondary structure representation of the “OFF” state. The segments within these two sequence highlighted in red display the secondary structure models for the AT helix as proposed in the literature.\textsuperscript{79,128} (c) & (d) P1 helix BPPs for transcripts with varying 3’ truncations for the two SAM-I riboswitches shown on the left. The horizontal axis plots the length increment with 0 starting at the 3’ end of the aptamer, the vertical axis shows the temperature (0-100 °C). The color scale represents the magnitude of the BPP as indicated in the legend on the right. Lines above the plots are color coded for the different structural elements (Terminator, AT and AT2) as indicated in the legend. These plots indicate that structures similar to the “OFF” state predominate until the transcript is long enough to incorporate the full AT, but the predicted behavior diverges for the two riboswitches for longer transcripts.
for non-canonical base pairing, is therefore less problematic than for some riboswitch systems. The simulations presented in this section only reflect the condition in the absence of the cognate ligand SAM.

Results from CTFSs of the two representative sequences, the *B. subtilis yitJ* and the *T. tengcongenensis metF* SAM-I riboswitches, are shown in Figure 3.2. The selected BPPs suggest the predicted population of P1 helix-forming riboswitches. Arguably formation of this helix makes formation of the terminator helix, and thus transcription attenuation, almost certain since the formation of the competing AT helix is blocked by P1 helix formation\textsuperscript{125,165}. Comparison of predictions reported in Figure 3.2 for the two sequences reveals contrasting patterns of transcript length dependent folding. For *B. subtilis yitJ* SAM-I riboswitch, P1 helix BPPs start to decrease significantly as the length of the RNA transcript reaches the position (length 32) that can form a full AT helix. P1 helix BPPs are restored when the 3’ strand of the rho-independent terminator has been “synthesized”.

The *T. tengcongenensis metF* SAM-I riboswitch displays a different pattern of transcript length dependent BPPs for the P1 helix. First, *metF* SAM-I riboswitch displays a significant temperature dependent behavior compared to *yitJ* SAM-I riboswitch. When the transcript is at the length that the AT1 helix model can form but not the AT2 helix model (Figure 3.3), there is less reduction of BPPs in the P1 helix at high temperatures (40-80 °C) in *metF* SAM-I riboswitch than in the *yitJ* SAM-I riboswitch. The model with the AT2 helix (Figure 3.3b) differs from the model originally suggested in ref\textsuperscript{158} (Figure 3.3a). In this AT2 model, the AT helix intrudes into the P4 helix as proposed in a subsequent biochemical study of this SAM-I riboswitch\textsuperscript{156}. A search for sequences with similar folding potential primarily yielded SAM-I riboswitches upstream of the *metF* gene from closely related thermophilic organisms (Figure 3.4 &
Figure 3.3 Two possible secondary structural modes of the antiterminator (AT) helix in the “ON” form for *T. tengcongensis* metF SAM-I riboswitch. (a) AT1 model as proposed in reference 157, in which the 5’ region can still form the secondary structure as observed in the crystal structure of the aptamer. (b) AT2 model as suggested in reference 156 and this study, in which the P4 helix of the aptamer is disrupted.

Figure 3.4 Sequence alignment of SAM-I riboswitch sequences that are upstream of the *metF* gene in several thermophilic anaerobic bacterium. Helical regions are masked in color blocks. Alternative secondary structural elements (P0 and AT2) are also shown in dot-bracket represent at corresponded regions.
APPENDIX A). The restoration for long transcripts is only predicted at low temperatures (0-40 °C). There is a wider span of lengths at which the P1 helix is disrupted in the *metF* SAM-I riboswitch (13 nts) than that in the *yitJ* SAM-I riboswitch (7 nts).

We wished to determine whether the differences between thermophilic and mesophilic SAM-I riboswitch folding patterns could be attributed to a specific modular component(s). The terminator sequence was swapped between these two sequences and then the same calculation repeated. Calculations in which the *yitJ* terminator replaces its counterpart in the *metF* SAM-I riboswitch are shown in Figure 3.5. The chimera *metF* now has a very similar transcript length dependent pattern as the original *yitJ* (Figure 3.2). The *yitJ* chimera still shows a pattern similar to that observed in the original *yitJ*. This sequence swapping experiment demonstrates that the SAM-I riboswitch transcription dependent behavior may be affected by the nature of the terminator element, if the system can reach an equilibrium distribution of conformation.

Prediction of thermodynamic parameters does not suggest that the *yitJ* terminator (-6.50 kcal/mol) is more stable than the *metF* terminator (-9.40 kcal/mol at 37 °C). Therefore, the restoration of the P1 helix when the terminator sequence is synthesized is not due to a more thermodynamically stable terminator. Further examination of secondary structure models suggests that the elimination of the possibility to form a stronger AT helix for *metF* causes this change (Figure 3.3b).

As demonstrated in the *in silico* sequence swapping experiment, the folding pattern changes dramatically when the natural balance among the P1 helix, the AT helix and the terminator is upset. Altogether, these results are consistent with the suggestions that variability in the expression platform plays an important role in tuning the function of riboswitches.\textsuperscript{160,168}
Figure 3.5 Sequence swapping test. The terminator sequences are exchanged between \textit{yitJ} and \textit{metF}, and BPP of the base pair highlighted with red arrow in the P1 helix is shown at the bottom of the sequence. (Left) \textit{yitJ} SAM-I riboswitch + \textit{metF} terminator, (Right) \textit{metF} SAM-I riboswitch + \textit{yitJ} terminator.

Figure 3.6 Heat maps of other \textit{B. subtilis} SAM-I riboswitch sequences from reference\textsuperscript{160}. The plot is same as Figure 3.2. Only the BPP of the base pair at the helical end close to the junction region is displayed here. Only the maximum length is displayed at the \textit{x} axis, the major unit of the \textit{x} axis is 5 nucleotides.
3.3.2 BPP Predictions for Additional SAM-I Riboswitches from *B. subtilis*

Calculations similar to those shown in Figure 3.2 were performed for a series of ten other SAM-I riboswitch sequences in *B. subtilis* characterized earlier by Tomsic et al\(^{160}\) (Figure 3.6). Further variability in predicted P1 BPP as a function of temperature and transcript length is observed amongst this group. Thus, tuning of riboswitch folding characteristics is predicted to vary amongst riboswitches within a single organism, as observed by Tomsic et al\(^{160}\), and is not solely due to the mesophilic or thermophilic origin of the riboswitch sequence.

3.3.3 Locating the Decision Windows in the SAM-I Riboswitch Sequences

The results reported in Figure 3.2 indicate that the transcript length at which either the P1 helix or the AT helix can form but the terminator is not yet fully transcribed could be an important decision point for SAM-I riboswitch regulated transcription. Here we present BPP calculations with this specific length of transcript. The lengths for *yitJ* and *metF* SAM-I riboswitches are 32 (151 nt in total) and 18 (128 nt in total) (as in Figure 3.2) beyond the aptamer, respectively.

As shown in the sequence diagrams of Figure 3.7, *yitJ* and *metF* SAM-I riboswitches have different competition topologies for the P1 and the AT helix. The *YitJ* SAM-I riboswitch has exactly the same base pair composition in either the P1 or the AT helix within the switching region. Six base pairs are involved in the pairing competition. The switching region is flanked by base pairs on both ends that are not involved in strand switching. For the *metF* SAM-I riboswitch, base pair composition (P1: 3 GCs, 7 AUs, 1 GU; AT1: 2 GCs, 5 AUs, 3 GUs, 1 UC mismatch) differs for the two alternative helices. In the model 1 AT helix, the P1/AT competition spreads through 11 base pairs. The AU base pairs at position 2 and 3 in the P1 helix have been shown to
Figure 3.7 Plots of BPPs for base pairs potentially involved in strand switching at selected temperatures at a fixed transcript length allowing for P1 or AT base pair formation. The schematic on the left illustrates the numbering scheme used to designate each switchable base pair position, as plotted on the horizontal axes. The “sum” is the summation of BPP values for the two possible base pairs each position. (a) BPPs for base pairs involved in formation of P1 or AT helix at 37 °C for the B.subtilis yitJ SAM-I riboswitch. (b) BPPs for base pairs involved in formation of P1 or AT helix at 37 °C and 65 °C for T.tengcongenesis metF SAM-I riboswitch. Note that for yitJ SAM-I riboswitch, positions one and two can only form P1 helix pairs, while only AT pairs can form for positions 9-13.
be important in the specific recognition of SAM via electrostatic interaction\textsuperscript{158,169}. In the \textit{yitJ} SAM-I riboswitch AT, the alternative base pair for this U is a closing a UA base pair in the AT helix, which should not be very stable. In two possible AT models of the \textit{metF} SAM-I riboswitch (Figure 3.3), the corresponding U either forms a UC mispair or unpaired two residue bulge. Therefore, this U appears to present a point of destabilization of the AT helix which may facilitate nucleation of a P1 helix. Also, in the case of the thermophilic SAM-I riboswitch, it may participate in a switch between alternative AT helices (see below).

In this BPP calculation, the BPPs for all base pairs in the P1 and the AT helix are displayed. The higher the BPP value, the more likely it is for the corresponding base pair to form at that position. If the summation of BPPs for two possible base pairs at one position is 1, an exclusive competition between P1 helix and AT helix base pairing is predicted for the corresponding position. The results confirm that both sequences favor formation of the AT helix at 37 °C (Figure 3.7). Nevertheless, the situation is reversed for \textit{metF} SAM-I riboswitch at 65 °C. The competition pattern for the P1 and the AT helix in the switching region also differs for the two sequences. Firstly, the exclusive competition region (sum > 0.8) in the \textit{yitJ} SAM-I riboswitch (position 3-8) is closer to the SAM binding pocket than that in the \textit{metF} SAM-I riboswitch (position 5-10). Secondly, the \textit{yitJ} SAM-I riboswitch displays a steady increase in the relative BPP values for AT versus P1 helix formation moving towards the right within the exclusive competition region, while the \textit{metF} SAM-I riboswitch has roughly the same relative BPP values for all positions. Interestingly, a hybrid structure with partial P1 and partial AT helix is observed in the minimal free energy (MFE) structure of the \textit{yitJ} SAM-I riboswitch (Figure 3.8). The hybrid structures observed here, as potential folding intermediates, are the basis for the RNA modeling in Chapter 4.
Figure 3.8 Secondary structure representation of possible hybrid intermediate states. These hybrid intermediate states are featured with partial P1 helix and partial AT helix. For example, the left one has 3 base pairs in the P1 and 10 base pairs in the AT helix, and the right one has 5 base pairs in the P1 and 8 base pairs in the AT helix. The hybrid geometry shown in the left figure is predicted in the MFE structure from RNAfold.

3.3.4 BPP Calculations Explain Reported Effects of Sequence Modifications on SAM-I Riboswitch Function: The Role of 5’ Residues in the “ON” State

Lu and colleagues\(^\text{159}\) reported that three mutations at the G11 residue within J1/2 of the \textit{yitJ} SAM-I riboswitch remove functional sensitivity to SAM. Paradoxically, termination for all three is enhanced even in the absence of SAM, though SAM binding is inhibited. The reported degree of enhancement from largest to smallest is G11C > G11U > G11A (Figure 3.9a). Since P1 helix formation is believed to be coupled to terminator formation, we consider it to be a predictor of transcription termination. In Figure 3.9b we calculate changes in the BPP for P1 helix formation relative to wild type for each of the three mutants in question, as well as relative to each other. At the chosen transcript length (which corresponds closely to the peak probability for AT helix formation) the mutations are predicted to increase the BPP for base pairs within the P1 helix, as observed\(^\text{159}\). G11C has the greatest increase, followed by G11U and G11A.
Figure 3.9 Effects of G11 mutation examined by BPP calculation. P1 helix BPP predictions for yitJ SAM-I riboswitch mutants at position G11 correlate, to some degree, with reported constitutive transcription termination. (a) Percentage of termination for SAM-I riboswitch without SAM and with SAM for different RNA constructs. The data is obtained from reference \(^8\). (b) The difference in BPP (P1 helix base pair at position 1) between two RNA constructs. The reference RNA construct is shown on the left of the plots, the color scale is as the color on the right. The darker the dot the greater the enhancement of the BPP in the specified mutant RNA construct compared with the reference construct. (c) (Top) Maximum BPP (BPMAX) for any one of the possible base pairs involving position 11 as a function of transcript length, (Middle) The sum of BPP (BPSUM) for all base pairs involving position 11, (Bottom) The total number of predicted base pair partners with position 11 (BPNUM). In each panel the predicted values are plotted for the WT and for each of the three possible mutations at the same position. Predicted dips in BPP for P1 helix formation (Figure 3.2, Figure 3.9b) broadly correlate with increased predicted base pairings involving position 11.
Therefore, the order of the rank from our BPP calculation is consistent with the experimental result. In Figure 3.9c, we observe that net probability of participation in any base pair (BPSUM) for residue 11 remains high for the mutants. The number of base pairing partners (BPNUM), on the other hand, is remarkably high for the WT sequence, and is reduced significantly for the mutants. The increase of base pair probability in the P1 helix in these mutants is therefore correlated with the elimination of alternative base pairing involving G11. P3 helix mutations which were reported to lead to constitutive transcription termination, however are predicted to diminish P1 helix BPPs or leave them unaffected.

3.3.5 Residues in J1/2 Are Predicted to Interact with Decoy Regions in the Absence of SAM

To detect where alternative interactions may be taking place within the BPP calculations, we monitored several possible alternative base pairings as a function of temperature for fixed lengths of the yitJ SAM-I riboswitch. Specifically we monitored base pairings involving residues in J1/2 and on the 5’ strand of the P1 helix. The choice of these monitors was based upon low energy secondary structures that could be observed in earlier simulations (data not shown). These alternative pairings are illustrated in Figure 3.10a. Figure 3.10b shows the selected BPPs as a function of temperature for a yitJ SAM-I riboswitch cut at length 151. As expected based upon Figure 3.2, base pairings in the AT helix predominate over P1 helix pairings, with the probability of the latter increasing with increasing temperature. Note that two “cross-junction” base pairings between the J1/2 and J3/4 appear with a high BPP at low temperatures. The appearance of these base pairs is inversely correlated with the increase in P1 helix BPP with temperature. However, monitors of P3 and P4 helix formation indicate only 60-80% BPP, suggesting predicted overall heterogeneity in secondary structure.
Figure 3.10. Decoys in the junction region. (a) Schematic illustration of alternative or “misfolded” conformations within the context of an “ON” state secondary structure previously proposed\cite{39,44,119} for the \textit{yitJ} SAM-I riboswitch from \textit{B subtilis}. Here D1 designates for decoy interaction 1—P0 helix, D2 for decoy 2—“Cross-Juction” and D3 for decoy 3—5’P1+J1/2 and 3’P4+J4/1. Predicted BPPs for individual base pairs for a 151 nucleotide length transcript are plotted as a function of temperature without (b) or with (c) constraints preventing participation of residues 87-90 in base pairing (e. g. pseudoknot formation). In part c, base pairing with two residues in J4/1 is also restricted, blocking the formation of “slipped” pairings with 5’ residues. BPPs are color coded according to the secondary structure element that each base pair participates in as indicated in (a). Restraints on base pairing increase the BPP for P1 helix formation at low temperatures, indicating that illustrated “decoy” base pairings favor AT formation.
The 11/87 and 12/86 cross-junction base pairings predicted in Figure 3.10b are inconsistent with the formation of a pseudoknot interaction between the apical loop of P2 and J3/4. The pseudoknot is observed in SAM-I riboswitch aptamer X-ray structures\textsuperscript{157,158,170}, and there is strong evidence for the functional requirement for this interaction in SAM-I riboswitches\textsuperscript{162,166}. RNAFold does not take account of the possibility of pseudoknot formation. Since formation of the pseudoknot would preclude other base pairings involving residues 87-90, we repeated the BPP calculations shown in Figure 3.10b with those residues constrained from base pairing altogether. The result shows that the overall competition trend between P1 and AT helices is little changed, and the probability of a fifth hairpin, termed P0, is increased slightly (data not shown). Slipped base pairs between 5-7 and J4/1 are represented prominently at low temperature. We repeated the calculations now constraining base pairing from residues 107 and 108 in J4/1, and pseudoknot residues in J3/4 as before. Figure 3.10c shows that with these constraints the temperature trend for the P1/AT helix competition is reversed, with higher temperatures weakly favoring AT helix formation. P0 helix formation is correlated with AT helix formation at low temperatures. The P0 helix is formed from a complementarity between residues in J1/2, including G11, and the 5’ region of the P1 helix. It is predicted to melt out above 20 degrees. In the higher temperature range, P1 and AT helix BPPs are relatively flat as a function of temperature.

The stabilization of the “ON” state by even a partial population of P0 helices could help to explain the G11 mutant observations in reference\textsuperscript{159}. We investigated whether such an additional helix could be predicted in other SAM-I riboswitches. Figure 3.11 shows that amongst the ~2900 SAM-I riboswitch sequences identified through Rfam\textsuperscript{171}, the vast majority (>95%) contain 3-5 potential base pairs between J1/2 and the 5’ strand of the P1 helix. Given the
constraints in the alignment (requiring the conserved AUC motif to sit in a loop region) the odds of identifying a stretch of three base pairs within random sequences in this location is less than 15%. For decoy interaction 3, sequence analysis shows that a subset of SAM-I riboswitches can form 8 or 9 base pairs in this region (Figure 3.11b). This decoy is not expected to be very stable because at least two bulges are embedded among base pairs.

Figure 3.11 Sequence survey for decoy interactions D1 and D2. (a) Statistics of the possible number of base pairs in the 5’ strand region for all the 2828 SAM-I riboswitches from Rfam. The sequences were truncated based on the sequence alignment to extract the strand that is composed of 5’ P1 and J1/2 segments as highlighted in the blue strand segment in the sequence schematic on the left. Each sequence was submitted for secondary structure prediction using RNAfold. The number of base pairs involved in the 5’ strand of the P1 helix were counted in the MFE structure. A back-of-envelope calculation shows that the odds for three consecutive base pairs to appear in the first 5 residues in a random sequence with 13 nucleotides with nucleotide 6-8 unpaired is ~14%. This number is significant smaller than 94.1% observed here. (b) Survey on the number of possible base pairs in the D3 region as shown in Figure 3.10a. The SAM-I riboswitch sequences used here are from the seed alignment from Rfam. The distribution of possible base pairs between these two regions in the SAM-I riboswitch sequences are compared to that of 10,000 random sequences generated by shuffling. The shuffling is only allowed inside individual strands for separated shuffling, while between two strands for single shuffling.
3.3.6 Incorporation of SAM Binding Effects into Secondary Structure Calculation

Above calculations, of course, have not taken account of the effect of ligand binding on SAM-I riboswitch secondary structure. It is well established that SAM binding favors formation of the “OFF” state secondary structure. From X-ray coordinates and studies of SAM binding to mutant riboswitches $^{157-159,165,172}$, we reasoned that we might derive some constraints on secondary structure base pairing as a consequence of SAM binding. We hypothesize that folding states that can bind SAM should satisfy the following minimal set of constraints: unpaired G11, no AU base pair in the AA-U internal loop in the P3 helix and 2 AU base pairs in the P1 helix. Here we calculate the BPP with the first and second constrained one at a time to test which constraint leads to the most significant enhancement of BPPs in the P1 helix (Figure 3.12).

![Figure 3.12 BPP calculation with constraints mimic SAM binding. (a) The difference between BPP (base pair at position 1 Figure 3.7) calculated with and without the specified constraints. Left: G11 unpaired, to mimic the effect of SAM contacting G11; Right: AA-U internal loop to mimic the effect of SAM replacing two As as potential pairing partners with U in P3 helix internal loop$^{158,173}$. (b) Comparison of BPPs for base pairs in the P1 and the AT helix at 37 °C with the constraint—G11 unpaired (solid line) and that without any constraint (dash). Within the region predicted to most readily favor AT helix formation, restricting G11 base pairing dramatically increases BPP for P1 helix formation.](image)
The constraint G11 unpaired causes the largest increase in BPPs in the P1 helix. The maximum difference is 0.663, which is observed at length 35 in the low temperature range. Figure 3.12b displays BPPs for all base pairs in the P1 and the AT helix at the same transcript length as in Figure 3.7a. The result indicates that the constraint G11 unpaired, which mimics one effect of SAM binding, increases the probability of P1 helix formation and reduces that of AT helix formation across the competing segment. This result is consistent with the prediction that G11 mutants possess less heterogeneous base pairing patterns for residue 11, since it is suggested that AT helix formation is correlated with conformational heterogeneity in 5’ residues. (Figure 3.9c & 3.10).

3.3.7 Experimental Probing of *T. tengcongensis* metF SAM-I Riboswitch Folding as a Function of Transcript Length

The BPP calculations described above appear to predict a number of features of full length *yitJ* SAM-I riboswitch conformational dynamics as reported in the literature. Aside from one recent study\(^{156}\), there is less data on secondary structure dynamics for *metF* SAM-I riboswitches. For both riboswitch sequences, most experimental reports have utilized transcripts which were truncated at or near the 3’ end of the putative aptamer, or which contained the full length riboswitch\(^{156,159,162,165,174}\). Here we have probed *T. tengcongensis* metF SAM-I riboswitch conformations for transcripts of varying lengths using several experimental methods.

We synthesized a series of three *metF* SAM-I riboswitch transcripts with 3’ cutoff points at the aptamer, at the AT1 helix, or at the proposed AT2 helix (Figure 3.13). In addition, to test for the possible impact of P0 helix formation on riboswitch conformation, we made similar length transcripts containing a pair of point mutations in J1/2 and in the 5’ strand of the P1 helix.
This pair of mutations was designed to enhance P0 helix formation by converting an AU pair to a GC, while destabilizing the competing P1 helix by disrupting a single base pair.

The folding of the three wild type RNAs is characterized using in-line probing (Figure 3.13). As expected, addition of SAM leads to a change in cleavage pattern for all three transcripts, which is similar to the SAM-induced pattern changes reported previously for full length yitJ riboswitches\textsuperscript{165}. The results are also consistent with other enzymatic and chemical probing reported for yitJ and metF SAM-I aptamers or full length riboswitches, in the presence and absence of SAM\textsuperscript{125,156,159,164}. In addition, the longest transcript probed (Figure 3.13b) shows a dramatic change in cleavage pattern in the P4 helix region as compared to shorter transcripts. The observed pattern is consistent with the formation of an AT2 helix, as predicted for transcripts of this length. Also, the AT2 model predicts that a GU dinucleotide (residues 99 and 100) is bulged out of the helix. Indeed, an enhanced cleavage in the absence of SAM is observed at these positions in Figure 3.13c as compared to Figure 3.13b. Significantly, the cleavage patterns in regions that are most sensitive to SAM binding, such as the pseudoknot, show an incomplete transition when SAM is added in Figure 3.13c. This observation is consistent with the hypothesis that AT2 helix formation hinders “OFF” state formation and SAM binding, as proposed in reference\textsuperscript{156} and as predicted above.

Figure 3.13a and Figure 3.13c also show that the cleavage pattern for the metF SAM-I riboswitch sequences is affected by magnesium concentrations in the absence of SAM. Some Mg-induced cleavage patterns, including that observed in residues in J3/4, parallel those associated with the addition of SAM (Figure 3.13a&c). This observation is consistent with fluorescence measurements that indicate that Mg enhances SAM-dependent folding of the yitJ SAM-I riboswitch\textsuperscript{163,174}. Figure 3.14 shows that introduction of the P0 helix stabilizing/P1 helix
Figure 3.13 Gel image from in-line probing experiment. In-line probing of secondary structure in metF SAM-I riboswitch-containing transcripts with varying 3’ truncation points. (a) Truncation at the aptamer, (b) Truncation at the “switch point”, at which residues comprising AT1 have been transcribed (c) Truncation at the point of complete transcription of AT2 helix. Red arrows: change of cleavage pattern in the P4 helix. Blue arrow: the increase cleavage is likely to be due to the GU bulge in the alternative AT helix model. J3/4 region is indicated as the arrow in magenta.
Figure 3.14 Gel image from in-line probing experiment. In-line probing experiment on wild-type (Left) and enhanced P0 mutant (Right). The area highlighted in red rectangle box is assigned to the GU highlighted in the previous figure. In the context of sequences truncated at AT1, they represent the fraying terminus of the AT1 helix. For the wild type sequence, this region has already been partially protected in the absence of SAM and fully protected in the presence of SAM. For the P0 mutant, the cleavage in this area is still persistent in the presence of low concentration SAM, indicating residual ON conformer formation.
destabilizing mutation, as predicted, appears to increase the residual “ON” state cleavage pattern when SAM is added. Gel mobility assays (Figure 3.15) also indicate that increasing transcript length to incorporate an AT2 helix, or introducing the P0 mutation, hinders formation of the SAM induced conformation.

![Native gel mobility of the T.tengcongenesis metF SAM-I riboswitch and mutants at different length of transcript. RNA was folded in 1x TB (pH 8.3), 200 mM KCl, 10 mM MgCl2, 20% Glycerol with or without 100 µM SAM. Sample A, WT aptamer (3’ end at position 0 as labeled in the T.tengcongenesis sequence in Fig. 1); sample B, WT AAT_AT_U2 (3’ end at position 16); sample C, eP0_AT_U2 (mutant to destabilize the P1 helix and stabilize the proposed P0 helix, 3’ end at position 16); sample D, 5’AUCP1_AT_U2 (WT AAT_AT_U2 with 5’ end sequence truncated); sample E, WT AAT_AT2 (3’ end at position 25); sample F, eP0 AT2 (P0 mutant 3’ end at position 25). A conformation state that has slow mobility and its population reduced in the presence of SAM is highlighted in red rectangle box. All samples show increased mobility with added SAM for the fastest running band, except D. The remaining samples also seem to indicate an increase in the relative population within the fastest running band, except F, which has been mutated to stabilize AT-forming conformation.]
3.3.8 Binding of SAM Is Reduced by the P0 and AT2 Helices

The inhibition of SAM-induced conformational change in in-line probing and gel mobility experiments by AT2 and putative P0 helices lead us to predict that these factors inhibit SAM binding by perturbing the conformational equilibrium towards the ON state. Equilibrium dialysis has been used as a means of direct measurement of the binding of SAM and other ligands to riboswitches \(^{165,175,176}\) (Boyapati et al, manuscript in preparation). We have utilized a \[^{3}\text{H}]-\text{SAM}\) as a reporter in single titration point equilibrium dialysis measurements, in order to rank binding affinity of the ligand to the RNA constructs shown in Figure 3.13. Binding seems to correlate with the experimental and predicted degree of “OFF” state formation. The strongest binding affinity appears with the shortest and intermediate length wild type constructs (Figure 3.16). RNA transcripts containing the P0 pair of mutations, still bind SAM but with lowered affinity. The wild type transcript long enough to form an AT2 helix, however, shows the weakest SAM binding affinity. Binding affinity is restored to P0 mutants by either lowering the temperature at which the dialysis incubation takes place (Figure 3.16b), or by introducing a compensatory mutation to stabilize P1 helix formation (Figure 3.16a).

3.3.9 5’ P1 Helix Residues Inhibit SAM-induced Displacement of a Fluorescent Reporter Oligonucleotide from \(T.\ tengcongensis\) metF SAM-1 Aptamer RNA

PyrrroloC is a fluorescent analog of cytidine that can base pair with guanine \(^{177}\). Upon incorporation within a base-paired A-form RNA helix, the fluorescence of pyrroloC decreases \(^{155}\). The reduced fluorescence signal is correlated with formation of hydrogen bonds as in a Watson-Crick GC base pair \(^{178}\). We designed an assay (Figure 3.17) to utilize this molecular probe to explore the conformational transition between the “ON” state and the “OFF” state. First, we
Figure 3.16. Effects of transcript length on SAM binding affinity examined by equilibrium dialysis. A higher b/a ratio indicates tighter binding of SAM. (a) Equilibrium dialysis (see Materials and Methods for details) performed with 3 µM RNA and 100 nM [3H] SAM. The annotation of different 3’ truncation points is the same as Figure 3.13a. WT denotes wild-type, eP0 designates a double mutant that enhances the stability of the P0 helix by converting an AU base pair to a CG base pair and introducing a GA mismatch at the P1 helix, and dP0 contains a single mutant that restores full P1 helix pairing and destabilizes the P0 helix by changing an AU base pair to CU mismatch. (b) Equilibrium dialysis performed with 3 µM RNA and 100 nM [3H] SAM at 4 °C and 22 °C. The annotation is the same as (a). (c) Equilibrium dialysis performed with RNA: [3H] SAM at 1:1 ratio at the concentration of 500 nM. RNA construct annotation is the same as (a).
Figure 3.17 Steady state fluorescence experiment on *T. tengcongensis* SAM-I riboswitch RNA constructs. (a) Titration of 3’ strand RNA of the P1 helix into solution containing Pyrrolo-C labeled 5’ strand RNA of the P1 helix. The fluorescence signal is recovered by adding excessive unlabeled 5’ strand RNA of the P1 helix. Schematic of the assay is shown at the left. (b) Equilibrium dialysis experiment to verify SAM binding of the RNA construct with a short P1 helix (8 base pairs). (c) Experiment to mimic the effects of SAM on shifting the conformational state towards the “OFF” form. Addition of 20 µM SAM to complex of reporter with aptamer containing partial P1 helix, results in partial recovery of fluorescence due to displacement of the reporter. (d) Control experiment showing that SAM does not displace the reporter from hybridization to a construct that cannot bind SAM due to truncation of 5’ residues, preventing P1 helix formation. (e) Same experiment as (c) but with an RNA construct with full wild type P1 helix.
hybridized single strand RNAs analogous to an isolated P1 helix to verify the capacity of pyrroloC as a conformation probe (Figure 3.17). We observed a dose-dependent reduction in fluorescence upon titration of a complementary unlabeled single strand RNA to a pyrroloC-containing analog of the 5’ strand of the P1 helix. Additionally, the fluorescence signal can be recovered upon addition of excess unlabeled competitor for the pyrroloC label RNA oligo.

Next, to mimic the effect of intramolecular interaction in a single sequence RNA, a 1:1 ratio of pyrroloC labeled RNA oligo and SAM-I riboswitch RNA constructs were hybridized. We performed the experiment with two SAM-I riboswitch aptamer RNA constructs. In the first construct, 4 nucleotides at the 5’ end were truncated, which results in only 8 base pairs in the P1 helix. In this case, the equilibrium should favor the hybridization of pyrroloC labeled RNA oligo to the 3’ strand region of the P1 helix. We used equilibrium dialysis to confirm the SAM binding capacity of this RNA construct (Figure 3.17b).

The fluorescence assay on this RNA indeed displays the maximum fluorescence reduction (~12% compare to ~9% in the full length P1 construct). Moreover, the fluorescence signal can be partially recovered in the presence of excessive SAM (10:1 ratio) (Figure 3.17c), while no effect from SAM has been observed in the control experiment (Figure 3.17d).

The same effect, however, is not observed in the experiment with a second aptamer construct with a full length P1 helix (Figure 3.17e). This finding leads us to suspect that the 5’ residues which were deleted in the first aptamer construct can participate in alternative base pairing. The lack of response to the presence of SAM in the full length P1 aptamer RNA construct may be due to formation of a more stable P0 helix at the 5’ region of the sequence, or to other “slipped” base pairings of 5’ pyrimidine residues with purines in J3/4, J4/1, or in the P4
helix region (Figure 3.10a). A UV melting experiment on an isolated sequence (Figure 3.18) also verifies the possible formation of a P0 helix.

![Diagram](image)

Figure 3.18 UV melting experiment of isolated P0 helix. (a) Sequence of the isolated P0 helix in secondary structure representation. (b) Melting profile of P0 helix in buffer condition: 50 mM KPhos, 150 mM KCl, 0.02 mM EDTA, pH=7.5 (similar condition as the buffer used in fluorescence experiment except without Mg2+). (c) Normalization of the melting data. (d) Thermodynamic parameters by fitting lnK vs. 1/T.

3.4 Discussions

We used BPP calculations, assuming a Boltzmann distribution of riboswitch conformers rather than a single lowest energy structure (MFE) or a two-state model, in order to understand
the equilibrium behavior of SAM-I riboswitches. Following from Quarta et al.’s\textsuperscript{161} study of the TPP riboswitch we analyzed the trend in BPP as a function of transcript length.

A striking observation from the BPP predictions is the link between alternative base pairing configurations for residues in J1/2, and the overall equilibrium between the two dominant conformers associated with “ON” and “OFF” states respectively (Figures 3.9 & 3.10). From these observations, we can hypothesize that SAM binding perturbs the equilibrium through contacts with J1/2, and G11 in particular, thus blocking the aforementioned alternative or “decoy” pairings. In this way, consideration of the conformational heterogeneity of the “ON” state seems crucial for a full understanding of the mechanism of coupling between SAM binding and riboswitch folding. In light of these predictions, previous suggestions that SAM binding may also indirectly immobilize J1/2 through magnesium contacts, gain added significance as a mechanism for ligand-induced folding \textsuperscript{179}. We have also identified positions at which folding switch points may be present, particularly if they correspond with transcriptional pauses. Certain predicted trends appear to be consistent with experimental probes of SAM-I riboswitch conformation reported here and elsewhere.

3.4.1 Base Pairing Patterns for the “ON” and “OFF” States of the SAM-I Riboswitch Differ at Three “Switch” Points

According to BPP predictions, at least two, and, for one riboswitch, three switch points can be identified during the course of transcription. At each point, at least one conformational decision takes place that potentially biases the final transcriptional state of the riboswitch. Figure 3.19 illustrates the three decision points schematically.
Figure 3.19 Schematic illustration of 3 conformational decision points during the synthesis of the T. tencongensis metF SAM-I riboswitch. The first decision point determines whether 5’ residues engage in distal interactions, leading to formation of a P1 helix, or whether they are sequestered by “decoy” purine residues in junction regions. The second decision point involves a competition between P1 and AT1 helix formation. The final decision point, which is not apparent in the B. subtilis yitJ SAM-I riboswitch, converts the AT1 helix to an AT2 helix. Up to this final switch leading to the formation of the AT2 helix and the transcription “ON” state (lower right), it is proposed that SAM binding could readily perturb the equilibrium towards the transcription “OFF” state (upper right).

First the P0 helix and “slipped” base pairings illustrated in Figure 3.10a may play a role in stabilizing the AT helix by occupying part of the switching region in the 5’ strand of the P1 helix. The “ON” state and “OFF” state secondary structures form the AT helix or the P1 helix, respectively. This choice represents the second structural switch as transcript length increases.
Since P0 helix and slipped base pair formation are incompatible with P1 formation, the first two switches are linked. Note that all of the pairings illustrated in Figure 3.10a, involving 5’ pyrimidine residues, could take place early in transcription and hinder P1 helix formation, and thus proceed to the “ON” state.

In the third potential switch point, predicted only in the case of \textit{T.tengcongenesis metF} SAM-I riboswitch, the AT helix can form two possible secondary structures. One, designated as AT1, resembling that proposed in reference \textsuperscript{157} may represent a switchable intermediate state (Figure 3.3). In this context, it is interesting to note that SAM can bind to \textit{yitJ} SAM-I riboswitch constructs which have been truncated so as to form hybrid P1/AT helices with \textmu M affinity (Boyapati et al, manuscript in preparation). The transcript length at which the BPP for P1 helix formation dips in Figure 3.7 ends with a stretch of Us for the \textit{metF} SAM-I riboswitch sequence. A stretch of Us has been shown likely to be a transcriptional pausing site \textsuperscript{180,181}. This stretch of Us has the potential to form a number of slipped base pairs resulting in either the AT1 model or the AT2 model in Figure 3.3.

The second AT helix model, resembling that shown in reference \textsuperscript{156} we call “AT2”. The AT2 conformer can only form for transcripts that extend into the terminator region. We identified several additional thermophilic SAM-I riboswitch sequences which align with the \textit{metF} SAM-I riboswitch sequence in a manner that allows for an AT2 helix model (Figure 3.3). Hennelly and Sanbonmatsu \textsuperscript{156} found that a fluorescent reporter hybridization designed to mimic an AT2 structure could not be displaced by SAM binding. We found a similar result with a fluorescent reporter hybridized to the 3’ P1 helix-forming strand of a SAM-I riboswitch aptamer. Yet in-line probing and equilibrium dialysis measurements indicate that AT2 forming transcripts do bind SAM at high concentrations. Thus our results suggest that the absolute irreversibility of
hybridization in fluorescent reporter assays is linked to the format. The suggestion, however, that formation of the AT2 helix represents a decision point hindering reversion to the “OFF” conformer even in the presence of SAM \(^{156}\) is consistent with our results.

3.4.2 Heterogeneous Models of the SAM-I Riboswitch “ON” State

Simulations presented in Figure 3.2 are consistent with the observation of a basal level of transcription termination without SAM in the *in vitro* transcriptional assay for the *B. subtilis yitJ* SAM-I riboswitch\(^ {159,160}\). Additional factors, however, such as alternative folding of the 5’ segment \(^ {182}\), kinetic trapping \(^ {183}\), tertiary contacts, other metabolites \(^ {184}\), etc., may be required for stabilizing the “ON” state conformation and to maintain expression of downstream genes in the absence of SAM.

Analysis of the current SAM-I riboswitch sequences in Rfam shows that at least three potential P0 helix base pairs are highly conserved. The addition of the P0 helix to the model for the “ON” state explains a number of otherwise curious observations. Firstly, the observation of constitutive transcription termination and “OFF” state secondary structure formation for the G11 mutants of the *yitJ* SAM-I riboswitch can be explained by the disruption of this helix. Some forms of alternative base pairings involving the 5’ residues can explain why SAM is able to compete off a fluorescent reporter oligonucleotide containing a 5’truncated aptamer (Figure 3.17c) but not from a full length aptamer (Figure 3.17e). Formation of the P0 helix may explain the observation in a previous study \(^ {156}\) that the J1/2 region is less solvent accessible in the absence of SAM.

It was not possible, however, to unambiguously obtain evidence for P0 helix formation, even in a mutant which was designed to stabilize the proposed helix. Altogether, the constrained
calculations show that at low temperatures, 5' P1 helix residues are predicted to be sequestered by “decoys” in J4/1, and possibly the P4 helix region, as well as the P0 helix. Given the limitations of the secondary structure predictions, none of the decoy interactions can be excluded from consideration. Indeed the entropic advantage of an ensemble of heterogeneous secondary structures could be a factor favoring “ON” state formation in the absence of SAM. A purine riboswitch variant which forms a native structure in the absence of ligand is apparently not found in nature. Nonetheless, cleavage of residues on one side of the P4 helix observed in in-line probing experiments by Winkler et al, are consistent with the presence of the D3 decoy interaction. Confirmation of this interpretation comes from in-line probing measurements on a yitJ hybrid riboswitch in which 5' P1 helix residues have been truncated, which do not show the P4 helix cleavage in the absence of SAM (Boyapati et al, manuscript in preparation). Functional SAM-I riboswitch sequences may need to maintain AT helix-stabilizing decoy interactions, just as the adenine riboswitch sequences are selected to allow antitermination.

3.4.3 Advantages and Limitations of BPP Calculations and the FEL Approach

Secondary structure prediction is the most popular computational application in RNA research. Multiple structures, such as suboptimal structures, or a Boltzmann weighted ensemble, can provide superior insight compared to a single secondary structure, especially for dynamic systems like riboswitches. Riboswitches typically act as “dimmer” switches allowing leaky expression, implying the simultaneous presence of permissive and non-permissive conformers. Overall, such behavior seems compatible with varying conformational population distributions, rather than a simple MFE or “all or none” two-state picture.
BPP predictions using a partition function, on the other hand, are limited by the free energy model, the choice of parameters, the form of the partition function and the neglect of tertiary interactions. The impact of these limitations, and, in particular, the consequence of not incorporating pseudoknot interactions, is discussed in the SI. Briefly, the BPP calculations are most valuable for exploring the functional role of competing conformers while details of predicted trends need to be interpreted cautiously. Nonetheless, several BPP predictions are compatible with some experimental observations: the observation of the AT1 to AT2 helix switch in the *metF* SAM-I riboswitch, the effect of mutations in J1/2 of the *yitJ* SAM-I riboswitch\(^\text{159}\), and the trend in the conformational equilibria with transcript length as reflected in the NMR spectra, gel mobility assays and in-line probing experiments reported here. Reports of constitutive termination in P3 mutants with reduced SAM binding\(^\text{159}\), however, do not correlate with the decreased predicted BPP for P1 helix formation. Tertiary interactions, or the impact of additional components of the transcription termination assay mix, appear necessary to explain those observations.

### 3.4.4 Implications for Folding of SAM-I Riboswitches in Active Transcription Complexes

Considering the general hierarchical folding mechanism and the constraints that secondary structures apply on the dynamics of RNA\(^\text{191}\), RNA folding at the secondary structure level can provide insights into the functional mechanisms of riboswitches. *In vitro* studies of SAM-I riboswitch structure, such as those implementing in-line probing or chemical and enzymatic reactivity, often incorporate long incubation times\(^\text{153,192}\). It is plausible to assume that the system reaches equilibrium under those conditions. For active transcription complexes, it cannot be assumed that RNA folding will reach its equilibrium state before a decision regarding transcription termination takes place. Predicting the folding behavior of such a system requires...
modeling rate constants for transcription, riboswitch folding, and ligand binding \(^{183,193-196}\). Moreover, transcriptional pausing will further complicate such predictions \(^{197}\). The presence of proteins in the cellular context, including the transcription complex in the 3’ region of the transcript, precludes a definitive set of predictions.

Nonetheless, our findings do have implications for modeling of SAM-I riboswitch-containing transcription complexes. Modeling of riboswitch folding kinetics requires consideration of folding pathways, which will be determined by the Free Energy Landscape. Our predictions and measurements indicate the key window during transcription for ligand binding or “ON” state folding to take place.

BPP calculations may also help to identify and predict differing functional properties observed for different riboswitches within a riboswitch class \(^{160}\). Arguably, calculations presented in Figure 3.2 predict that a full length \(metF\) SAM-I riboswitch sequence may manifest less “leaky” transcription termination in the absence of SAM than the \(yitJ\) SAM-I sequence. We note that methylation of lysine accounts for the increased thermostability for the proteins in thermophile archaea \(^{198}\). As a result, thermophiles may require a tighter control of the SAM pool. Interestingly, the BPPs for P1 helix base pairing increase in the temperature range around the optimal growth temperature (65 °C) \(^{199}\) of this thermophillic species. The function of the extended AT helix in stabilizing an ON conformation in the thermophilic riboswitch may be linked to the need for the riboswitch to function in a higher temperature range.
CHAPTER 4 A ROLE OF SMALL MOLECULE IN STRAND SWITCHING FROM RNA MODELING AND MD SIMULATION
4.1 Introduction

Most riboswitch structural and biophysical studies until now have conceptualized the problem as recognition of a small molecule by a folded RNA secondary structure termed the “aptamer”. A number of X-ray and NMR structures of aptamer-ligand complexes have helped to explain sequence conservation patterns observed in aptamer domains \(^{71-79,82,87}\). On the other hand, a complete description of the switching mechanism also demands insight into the structure and dynamics of the so-called “expression platform”. The expression platform is located immediately downstream of the aptamer and directly controls gene expression. In fact, the crucial determinant of riboswitch function is usually linked to a choice between alternative base pairing partners for a “switching” RNA segment. The result is formation one of two competing helices, the Anti-Terminator (AT) or the Terminator (T) in the expression region. Moreover, solution and biophysical studies of unliganded riboswitch domains indicate a more dynamic picture than the ligand-bound X-ray structures, indicating dynamical nature of the switching mechanism \(^{156,168,199-205}\). These considerations raise questions as to how riboswitch ligand binding discriminates between alternative base pairing configurations, and the direct effects of ligand binding on the equilibrium between such configurations. However, our understanding of the switching mechanism has been hampered by the lack of 3D models for the SAM-I riboswitch including the expression platform.

We recently observed binding of SAM to hybrid constructs derived from the SAM-I riboswitch (Boyapati et al, manuscript in preparation) (Figure 4.1). These hybrids encompass the AT element and the sequence of the core tertiary structure-forming region of the aptamer domain. In addition, they contain a varying number of 5’-end nucleotides from the P1 helix, indicating that SAM could bind even when the AT helix is formed but the P1 stem is not complete. Indeed,
several studies suggest that P1 helix formation is not stable in the absence of SAM $^{87,173}$. Therefore, during an *in vitro* or *in vivo* transcription the AT element may form a local hairpin structure before the P1 stem formation. Here we raise the question as to whether SAM binding to such intermediates could trigger a potential branch migration leading to full P1 stem formation thus completing aptamer formation. While many studies suggest the importance of the P1 stem formation for the SAM-binding assisted switching mechanism with SAM-I riboswitches $^{174}$, such an interplay between the expression platform and the aptamer domain has never drawn attention in the literature.

<table>
<thead>
<tr>
<th>b/a ratio</th>
<th>1.3±0.1</th>
<th>4.4±1.3</th>
<th>1.2±0.1</th>
</tr>
</thead>
</table>

*Figure 4.1 Schematic of observations from SAM binding experiment. A higher b/a ratio indicates tighter binding of SAM, b/a ratio close to 1 means no binding or weak binding.*

RNA 3D structures are generally regarded as difficult to model, partly due to their flexibility in the backbone $^{206}$ as compared to many classes of proteins. Riboswitches may be good model systems for conformational changes in response to binding of metabolites. Recent advances in theoretical and software development have enabled the building of RNA models.
semi-automatically \textsuperscript{207} or automatically \textsuperscript{208-211}, and in fact notable achievements were reported \textsuperscript{212}. On the other hand, all-atom MD simulations are powerful computational approaches for atomistic details but generally limited to time scales less than microseconds. The birth of a specialized machine designed for MD simulation—Anton \textsuperscript{213} has fueled the rocket to escape the timescale limitation. In addition, progress in implementing various enhance sampling techniques \textsuperscript{214-216} also equip us with alternative tools for exploring the conformational space of large biomolecule systems.

Taken together, the objectives in this study are twofold. First, we wish to evaluate a strategy combining 3D modeling with all-atom multiple MD simulations to probe RNA conformational dynamics that require an exceptionally long timescale. Another aim is to investigate the possible presence of a branch migration mechanism that explains the interplay between the aptamer formation and ligand binding.

To this end, we employed the combination of RNA 3D modeling with multiple all-atom MD simulations. A model riboswitch segment is designed in a manner which allows the survey is focused on a limited portion of the energy landscape (EL) \textsuperscript{217-221} providing a clue for the possibility of the branch migration. In this study, starting with a model with partial P1 helix, we observed formation of a complete P1 helix through invasion of the base pairing region of the AT helix in the presence of SAM. Additionally, we depicted a possible free energy landscape of a SAM-I riboswitch by combining discrete conformational space sampling using a modeling package—MC-sym with all-atom MD simulations. Our study synergizes computational techniques with experimental inputs to tackle an important biological problem. The results from this study suggest a plausible pathway for the strand switching.
4.2 Structure Modeling Using MC-sym

4.2.1 RNA Modeling Using MC-sym

To understand the potential role of branch migration process in SAM-I riboswitch mechanism, we use the RNA construct—6P1_11AT as a model system (Figure 3.2). The branch migration process is hypothesized such that the competition of two helices—the P1 and the Anti-Terminator (AT) helix determines the fate of two alternative structures for the ON state and the OFF state even when the 3’ strand of the AT helix is synthesized. In other words, the 5’ nucleotides of the P1 helix may invade the formed AT helix and eventually form a complete P1 helix, and thus this mechanism contributes as one of folding pathways of SAM-binding assisted aptamer domain formation and thus the Terminator (T) formation, i.e. the OFF state. Here are several arguments for motivating such a hypothesis for a role of this pathway. First, considering the general hierarchical folding mechanism for RNA, the local helices such as AT helix, along with P2, P3, and P4 would form more rapidly than the P1 helix which requires the long distance interaction to close the four-way junction structure comprising P1, P2, P3, and P4. In addition, a number of decoy base pairings which could further slow down the folding process have been described in Chapter 3. Secondly, it is accepted as a consensus that SAM binding could facilitate the tertiary structure formation associated with inter-helix interactions such as the pseudoknot formation between P2 and J3/4 and the kink-turn formation, which helps the formation of the P1 helix by bringing together two distal strand segments. Then, the question is that when the concentration of metabolite is low, the P1 helix formation cannot form while the AT is formed. Thirdly, a SAM binding assay with various constructs that have a variable cut-off at 5’ end of the P1 region while the 3’ end is long enough to allow the AT helix formation, found that still SAM binding was detected (Boyaptai et al, manuscript in preparation). The observations suggest that
before the full P1 formation, with the AT helix, SAM could be bound. Taken together, these considerations raise the question as to whether SAM binding would then drive a conversion to the “OFF” state containing a fully formed P1 helix.

Figure 4.2 Secondary structure representation of the hybrid construct--6P1_11AT. This hybrid system contains the AT helix (in red) that is assumed to be formed and two base pairs in the P1 helix (in blue) act as the nucleation site of the P1 helix. The remaining P2, P3, and P4 helices and tertiary structures including a pseudoknot structure and a kink-turn motif (in blue) are assumed to form and derived without change from the AAT X-ray structure. The base triplet A4-U110-A136 (in purple) that is one set of nucleotides that involves in the P1 helix and the AT helix competition. Three additional nucleotides (in green) in the 5’ strand of the P1 helix are modeled as single strand RNA. The same color scheme is used for the structure view in this chapter except otherwise mentioned.

The atomic models for the RNA construct as shown in Figure 4.2 were generated using MC-sym installed locally. The blue part of this construct was modeled using its counterpart in the X-ray coordinates of the yitJ SAM-I riboswitch (PDB ID: 3NPB). The other parts of the construct were built from the library of small fragment RNA structures, known as Nucleotide
Cyclic Motifs (NCMs)\(^{222}\). An explicit triplet constraint was applied on the three nucleotides highlighted in red (A1/4, U107/110 and A133/136) to allow sampling the 3D space in which these three nucleotides are in vicinity to each other. In these three nucleotides, the two As are competing for base pairing with a U. MC-sym uses a RMSD cutoff value to determine whether a new model would be accepted. Here different RMSD threshold values are tested in the sampling procedure using MC-sym to ensure the exhausted sampling in the local region bridging the partial P1 and the AT helix (A109, U110 and A111). Energy minimizations (max step is 2000 or rms of energy < 1.0) were performed on the atomic structures of the models generated from MC-sym runs using Nucleic Acid Builder (NAB)\(^{223}\). AMBER99bsc0 force field\(^{224}\) and Generalized Born model\(^{225}\) with an inverse Debye-Hückel length of 0.19 Å\(^{-1}\) were used in the energy minimization procedure. 149 models were generated in this step. This energy minimization protocol is mainly to rebuild the chain connectivity for models generated from MC-sym without introducing the sampling effect of the force field. Therefore, here we used MC-sym to sample the possible placement of the AT helix in the 3D structures and the geometry of the potential nucleation site of the P1 helix close to the SAM binding pocket. The modeling is based on the assumption that the folding of the remaining coordinates is similar to that in the crystal structure of the aptamer domain. After the energy minimization step, models with high van der Waals energy were filtered out. There are two reasons for high van der Waals energy: 1) steric clashes that cannot be released by energy minimization, 2) broken chain connectivity that cannot be bridged during energy minimization.

4.2.2 All-atom Molecular Dynamics Simulation

Models were chosen following these criteria: 1) Top rank models in calculated free
energy using Amber99bsc0 force field with GB implicit solvent model. 2) The SAM binding pocket is not occupied. 3) Coaxial stacking of the AT helix and the P1 helix.

Models satisfying the criteria above were chosen for MD simulations. For the models in the presence of SAM, the ligand was placed in the binding pocket maintain most of the interactions (except the contacts with the end base pair AU in the partial P1 helix) observed in the crystal structure of the aptamer domain complex (PDB: 3NPB). For short trajectories (< 100 ns), the simulations are performed using NAMD 2.8 227 with Amber99bsc0 force field 224 on Teragrid machines. For long trajectories (> 500 ns), the simulations are run on Anton 213. Parameters for SAM are from the Generalized Amber Force Field (GAFF 228) and partial charges are calculated using ANTECHAMBER 229 using the AM1-bcc model 230. The starting structures are prepared using the LEaP module in AMBER 231. The RNA molecules are solvated in a cubic solvent box of TIP3P waters with a 12 Å (10 Å for Anton simulations) padding in all directions. Sodium ions are placed in the most negative position around RNA in order to neutralize charges of the system. Energy minimizations are carried out to remove bad contacts until the RMSG < 1.0. Starting from 0 K, the temperature is raised 10 K for every 10 000 steps and is held constant after reaching the desired temperature (300 K) using temperature reassignment. Harmonic constraints were applied on the RNA and gradually removed over the equilibration steps using scale from 1.0 to 0 with 0.2 increment per 50 000 steps. MD simulations are performed in the NPT ensemble. The Langevin dynamics was used for constant temperature control, with the value of Langevin coupling coefficient and the Langevin temperature set to 5 ps and 300 K, respectively. The pressure was maintained at 1 atm using the Langevin piston method with a period of 100 fs and decay times of 50 fs. The time step is 2 fs for both equilibration and production phase. Bond lengths between hydrogens and heavy atoms are constrained using
4.2 Structure Modeling Using MC-sym

4.2.1 RNA Modeling Using MC-sym

To understand the potential role of branch migration process in SAM-I riboswitch mechanism, we use the RNA construct—6P1_11AT as a model system (Figure 3.2). The branch migration process is hypothesized such that the competition of two helices—the P1 and the Anti-Terminator (AT) helix determines the fate of two alternative structures for the ON state and the OFF state even when the 3’ strand of the AT helix is synthesized. In other words, the 5’ nucleotides of the P1 helix may invade the formed AT helix and eventually form a complete P1 helix, and thus this mechanism contributes as one of folding pathways of SAM-binding assisted aptamer domain formation and thus the Terminator (T) formation, i.e. the OFF state. Here are several arguments for motivating such a hypothesis for a role of this pathway. First, considering the general hierarchical folding mechanism for RNA, the local helices such as AT helix, along with P2, P3, and P4 would form more rapidly than the P1 helix which requires the long distance interaction to close the four-way junction structure comprising P1, P2, P3, and P4. In addition, a number of decoy base pairings which could further slow down the folding process have been described in Chapter 3. Secondly, it is accepted as a consensus that SAM binding could facilitate the tertiary structure formation associated with inter-helix interactions such as the pseudoknot formation between P2 and J3/4 and the kink-turn formation, which helps the formation of the P1 helix by bringing together two distal strand segments. Then, the question is that when the concentration of metabolite is low, the P1 helix formation cannot form while the AT is formed. Thirdly, a SAM binding assay with various constructs that have a variable cut-off at 5’ end of the P1 region while the 3’ end is long enough to allow the AT helix formation, found that still SAM binding was detected (Boyaptai et al, manuscript in preparation). The observations suggest that
Figure 4.3 MC-sym sampling benchmark. (a) The number of models that MC-sym generated using different RMSD threshold within the same amount of CPU time. (b) Histogram of the pseudo-dihedral (Figure 4.4) between A109 and U110, the single strand region that connect the partial P1 helix and the AT helix in 6P1_11AT, in models sampled from different RMSD threshold values. Here pseudo-dihedral was chosen because it is a generalized order parameter to describe the geometry of neighboring nucleotides.

The hybrid constructs, and that SAM binding has similar dependence on Mg\(^{2+}\), and similar sensitivity to mutations as with the aptamer (Boyapati et al, manuscript in preparation), indicates that this assumption is approximately correct when SAM is present. Final structural models from MC-sym are shown in Figure 4.4b. The results demonstrate that MC-sym can sample a wide range of the conformational space to arrange the AT helix without steric clash with the other part of the structure. One of the critical local region to be sampled in the particular model is the three nucleotide A109, U110 and A111. These three nucleotides act as a hinge to bridge the partial P1 helix and the incomplete AT helix. Monitor of the pesudo-dihedral angle (Figure 4.4c) also displays that MC-sym has sampled exhaustedly the possible local geometry from known structures without biasing to highly populated geometry observed from crystal structures. Noticeably, there is a region (between 80 and 170 degree) that is rarely sampled. This is because
Figure 4.4 Structural analysis of MC-sym models. (a) Superposition of all the models from MC-sym in tube representation. The AT helix is shown in red, and the other part of the RNA is shown in blue. (b) Schematic of the pseudo-dihedral angle definition from this reference. (c) (Top) Histogram of pseudo-dihedral angles for 5’-AU-3’ dinucleotide (Left) and 5’-UA-3’ dinucleotide (Right) in known RNA structures. (Bottom) Histogram of pseudo-dihedral angles for A109-U110 dinucleotide and U110-A111 dinucleotide sampled by MC-sym.
Figure 4.5 Structure analysis of MC-sym models. (a) Mapping vdW interaction energies vdW_{A109-U110} (Left) and vdW_{U110-A111} (Right) on the coordinate set describes the relative nucleobase orientation in the consecutive triple nucleotides (A109, U110 and A111) with two pseudo-dihedrals—ϕ_{A109-U110} and ϕ_{U110-A111}. The open circles are the discrete models generated from MC-sym. The surface of the vdW interaction energy is plotted using grid fitting and the nearest data point determines the color on the map. The colors for regions without data points are predicted using the two flanking data points, and thus do not necessarily reflect the actual energy surface. (b) Scatter plot of vdW_{A109-U110} and vdW_{U110-A111} clustered into 4 classes and the global view of model 51 and model 55 in cartoon representation (c) Plot of the free energy rank with amber99bsc0 and GB solvent model. (d) Schematic of the local switching region. The two arrows indicate the order for numbering the base pairs. (e) Local 3D view of model 51 and model 55 in the region as shown in (d).
the closeness of the P1 helix to the P3 helix in the crystal structure places some steric restrictions on assembling fragments with $\phi$ in the range of 80-170° into the models by MC-sym.

One of the significant observations from the equilibrium dialysis experiments (Boyaptai et al, manuscript in preparation) is the possible co-axial stack between the P1 and the AT helix. The three nucleotides A109, U110 and A111 should play an important role in constructing the continuous co-axial stack. Here the vdW interaction energy between the nucleobases of the adjacent nucleotide is employed to examine the base stacking. Figure 4.5a shows the vdW energies of the stacking for two sets neighbor nucleotides A109-U110 (Left) and U110-A111 (Right) plotted on the same coordinate reference—the pseudo-dihedral angle of A109-U110 as the $x$ axis and the pseudo-dihedral angle of U110-A111 as the $y$ axis. The results show only a limited conformational space is allowed for continuous co-axial stacking through these three nucleotides. This conformational space is indicated by the common area with negative vdW interaction energy in both plots (Figure 4.5a). These structural models from MC-sym can also be classified into four clusters, based on squared Euclidean distance on the 2D space using $k$-means ($k=4$) clustering algorithm, as shown in Figure 4.5b. Cluster 1 has no or very weak stacking between A109 and U110 and cluster 4 has no or very weak stacking between U110 and A111. One structural feature of these two clusters is that AT helix adopt a wide range of angle relative to the P1 helix, and thus in most cases the co-axial stacking is unfavorable. This structural feature also appears in cluster 3, in which strong stacking energies are not observed in either adjacent dinucleotide. Models with good co-axial stacking should locate in cluster 2 or be some very close neighbors of cluster 2. Additionally, these structure models are also evaluated using amber99bsc0 force field (Figure 4.5c). Two (model 51 and model 55) out of the top five ranks in terms of free energy by force field satisfied all the three criteria as described in the method.
The global of these two structures are shown in Figure 4.5b, and the local geometry of the switching region is displayed in Figure 4.5d. The main difference between these two models is that the AT helix adopts a different twist angle relative to the P4/P1 helix, and thus the 5’ strand of the P1 helix is placed in the two different grooves of the AT helix—model 51 in the minor groove of the AT helix, while model 55 in the major groove. The geometry sampled in these two models resembles RNA triple helix composed of poly(U)-poly(A)-poly(U) from a crystal structure, but without an undistorted groove region in our models. With limited experimental data, these two models are rationalized as the potential models for the intermediate or “transient state” between “ON” and “OFF” state. Therefore, these two models are submitted for MD simulations in the absence and in the presence of SAM to explore their motions in dynamics. Meanwhile, MD simulations are also performed on the crystal structure of this aptamer domain from yitJ SAM-I riboswitch for reference purpose.

4.3.3 Transient Ensemble by MD Simulations

Table 4.1 lists the trajectories included in this study. Different trajectory evolutions are observed for model 51 with or without SAM. Strikingly, formation of a complete P1 helix (all 6 Watson-Crick base pair) is presented in the trajectory at ~1.3 µs for the simulation in the presence of SAM. Figure 4.6 displays the time evolution of

<table>
<thead>
<tr>
<th>Models</th>
<th>SAM</th>
<th>Traj Length (µs)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Model 51</td>
<td>+</td>
<td>3.128</td>
</tr>
<tr>
<td>Model 51 (frame6615)</td>
<td>+</td>
<td>0.884</td>
</tr>
<tr>
<td>Model 51</td>
<td></td>
<td>1.256</td>
</tr>
<tr>
<td>Model 55</td>
<td>+</td>
<td>1.467</td>
</tr>
<tr>
<td>Model 55</td>
<td></td>
<td>0.642</td>
</tr>
<tr>
<td>3NBP</td>
<td>+</td>
<td>1.280</td>
</tr>
<tr>
<td>3NBP</td>
<td></td>
<td>0.870</td>
</tr>
</tbody>
</table>
RMSD for individual base pairs with reference to that in the P1 helix of the crystal structure. A smaller RMSD value (deep blue) indicates the geometry of the nucleobases in a single base pair is closer to the Watson-Crick base pair observed in the crystal structure. The monitor of hydrogen bond presence for these hydrogen bonds in Watson-Crick base pairs for the base pairs in the P1 helix and the AT helix are presented in Figure 4.7. Combining information from Figure 4.6 and Figure 4.7, the time that a completely formed P1 helix can be located as indicated with the red arrow in Figure 4.6, and the lifetime of this conformation span from frame 6544 to frame 6635 (18.2 ns). Fraying of the closing base pair (base pair 6) in the P1 helix explains the short life span of this conformation, and these two participating nucleotides flip to a dominant cross-strand stacking conformation after this closing base pair is lost. The second closing base pair (base pair 5) in the P1 helix also has a small population in the Watson-Crick base pair state (2250 out of 15641 frames). This base pair is disrupted shortly after the loss of the closing base pair, but reappears at ~1.9 μs. However, the other 4 base pairs in the P1 helix (base pair 1 to 4) maintain the P1-like conformation in the crystal structure during this simulation in the presence of SAM. Additionally, the electrostatic interactions between the sulfur atom of SAM and the carbonyl oxygen atoms of two U residues persist through out the simulation as observed in the repeated simulation on the aptamer domain of yitJ SAM-I riboswitch (Figure 4.8). The same analysis on the AT helix shows that the destabilization of AT base pair 1 to 3 precedes the complete P1 formation. The disruption of this AT region is not due to the deficiency in modeling the AT helix since the simulation on the same model without SAM maintains the geometry close to a standard A-form helix for 2 (base pair 2 and 3) out of these three base pairs. Moreover, there is no sign of the 5’ P1 strand taking over the switching region for model 51
Figure 4.6 Time evolution of RMSD for individual base pairs in the P1 helix and the AT helix. Time evolution of RMSD for individual base pairs in the P1 and the AT helix from simulations on model 51 and model 55 in the presence and in the absence of SAM. The crystal structure of the P1 helix from yatJ was used as the reference structure for the P1 helix, while a standard A-form helix was used as reference for the AT helix. The starting structure and some snapshots from the trajectories were shown around the RMSD plot. The red arrow highlights the time period that the complete P1 is sampled in model 51 with SAM.
Figure 4.7 Monitor of hydrogen bonds in Watson-Crick base pairs. Here hydrogen bonds are defined with H-bond distance cutoff (< 3.5 Å) and H-bond angle cutoff (> 145°). The order of hydrogen bonds in the P1 helix is as the following: 1: C6@O2--G108@N2, 2: C6@N3--G108@N1, 3: C6@N4--G108@O6, 4: U5@N3--A109@N1, 5: U5@O4--A109@N6, 6: A4@N6--U110@O4, 7: A4@N1--U110@N3, 8: U3@N3--A111@N1, 9: U3@O4--A111@N6, 10: U2@N3--A112@N1, 11: U2@O4--A112@N6, 12: C1@O2--G113@N2, 13: C1@N3--G113@N1, 14: C1@N4--G113@O6. The list of hydrogen bond in the AT helix is as the following: 1: A136@N6--U110@O4; 2: A136@N1--U110@N3; 3: U135@N3--A111@N1, 4: U135@O4--A111@N6, 5: U134@N3--A112@N1, 6: U134@O4--A112@N6, 7: C133@O2--G113@N2, 8: C133@N3--G113@N1, 9: C133@N4--G113@O6, 10: U132@N3--A114@N1, 11: U132@O4--A114@N6, 12: U131@N3--A115@N1, 13: U131@O4--A115@N6, 14: C130@O2--G116@N2, 15: C130@N3--G116@N1, 16: C130@N4--G116@O6.
Figure 4.8 RNA-ligand interaction distance monitor. (Left) Highlight of the interaction between SAM and residues in the binding pocket observed in the crystal structure. The adenine moiety of SAM stacks on residue C46. The adenine ring and methionine moiety of SAM form a network of hydrogen bonds with residues in J1/2 and P3 (G11, A44, G76 and G77). The readout of positively charged sulfur on SAM is achieved by two carbonyl oxygen atoms (O2) on U5 and U110. The numbering of residues follows that in 6P1_11AT. (Right) Monitor of the distances for the electrostatic interactions between RNA and SAM during the simulation for model 51 and 3NPB in the presence of SAM.

Figure 4.9 Radius of gyration of different SAM-I riboswitch models during simulations.
without SAM. On the other hand, the partial P1 helix with contacts to SAM exposes its instability in the absence of SAM.

Although the effect of SAM on maintaining the compactness of the RNA is demonstrated in both modeled systems (Figure 4.9), the complete P1 formation does not take place in model 55 within the time scale (1.467 µs) accessible so far for this simulation when SAM is present. Base pair 3 in the P1 helix does not sample a Watson-Crick base pair, but instead traps in a state with base pair geometry close to an AU Hoogsteen base pair ($U\cdot A$ cis W.C./Hoogsteen and class XXIII according to reference $^{234}$) (Figure 4.10). In addition, this state is stabilized by a new hydrogen bond interaction between A4 and SAM (Figure 4.10), which is not sampled during the simulation of the aptamer domain in the presence of SAM. Furthermore, this AU Hoogsteen base pair enables the co-axial stacking between the partial P1 helix and the AT helix through the
stacking between A4 and A111 (Figure 4.10). This cross-strand stacking competes over the relatively weak (-1.99 kcal/mol) adjacent stacking between U110 and A111 in the starting structure—model 55. Further examination of structural models from MC-sym in Cluster 2 (Figure 4.5) shows that models with $E_{vdW} > -2$ kcal/mol for the U110-A111 adjacent stacking do not allow the placement of the 5’ strand of P1 helix in the major groove region of the AT helix as that in model 55. However, models similar to model 51, with the 5’ strand of P1 in the minor groove of the AT helix, dominate in Cluster 2, especially in the region both $E_{vdW} > -4$ kcal/mol (Figure 4.5).

4.3.4 Different Modes of Motion in the P1 Helix

We now employ PCA (as described in Chapter 2) to characterize large conformational change from multiple trajectories. Unlike previous PCA, the analysis focuses on the dynamics of local structural elements—the P1 helix. Because all the systems as listed in Table 4.1 include the section of the P1 helix incorporating (residue 1-6, 108-113 for model 51&55, residue 4-9, 111-116 for 3NPB), the conformation sampled for this part of P1 helix from all the six trajectories are used as inputs for PCA so that the trajectories can be visualized within the same reference coordinate set in the conformational space with reduced dimensionality.

Again, fluctuations along the first several eigenvectors contribute to a large portion of overall dynamics in the P1 helix (Figure 4.11a). 2D projection of trajectories onto the first two eigenvectors—the first two principle components are used to visualize the conformational sampling from these multiple trajectories with different starting structures in a reduced conformational space (Figure 4.11b). Also, the same 2D projections of individual trajectories are
Figure 4.11 PCA of the P1 helix from multiple starting structures. (a) Eigenvalue and percentage of fluctuation for the first 20 eigenvector from PCA of mixed trajectories, including model 51 + SAM, model 51 – SAM, model 55 + SAM, model 55 – SAM, 3NPB + SAM and 3NPB – SAM. Only the residues present in all systems are extracted for this analysis—residue 1-6, 108-113 for model 51 & 55, residue 4-9, 111-116 for 3NPB. (b) Smoothed 2D density plot for the 2D projection of the mixed trajectories onto the first two eigenvectors—principle component 1 (PC1) and PC2. The interval between bins is 0.1. The color scale is as color bar on the right. Regions with density less than 3 are plotted using original data points as white dots. (c) Visualization of motions along the first two eigenvectors using porcupine plots.
Figure 4.12 2D projection of individual trajectory onto the first two eigenvectors for PCA of the P1 helix. Trajectories with SAM are shown in red, and without SAM in blue. The probability density plot for PC1 is displayed at the top of each panel, and PC2 along the right axis of the panel. Probability density plots for the same PC have the same vertical scale—0.13 for PC1 and 0.2 for PC2.

displayed in Figure 4.12. In addition, the large-scale concerted motions along the first two eigenvectors are visualized in porcupine plots (as described in Chapter 2). The porcupine-plot along the 1st eigenvector displays a large swing motion of three nucleotides at the 5’ end of the P1 helix. This motion captured in the 1st eigenvector reflects the distinctions among starting structures obtained from MC-sym modeling and the crystal structure. The starting points of the three arrows feature the placement of the 5’ dangling end nucleotides in model 55, the structure in the figure is the representative of this P1 part in the crystal structure, and the end points reflect the accommodation of these three nucleotide in model 51. This result showcases the capacity of discrete conformational space sampling using MC-sym to escape the time-scale limitation even with MD simulations in µs time scale. Interestingly, the effects of SAM are presented in the
motion along the 2nd eigenvector. A major motion mode is the slippage of base pair 1-3 as indicated by the arrows in opposite directions for the three nucleotides on two different strands. Two (base pair 1 and 2) out of these three base pairs are modeled as Watson-Crick base pairs in model 51 and model 55. In the presence of SAM, the persistent electrostatic interactions between SAM and two Us on two different stands during the simulation (Figure 4.8) lock this local geometry in a well-stacked base pairing mode (Figure 4.12). On the other hand, this part of the modeled structure became unstable and fell apart in the absence of SAM for model 51. The disruption of this partial P1 helix without SAM is not observed in model 55 and 3NPB. For model 55, this is because A136 intercalates at the 3’ end boundary between the partial P1 helix and the AT helix, and thus mediates a co-axial stacking to stabilize partially formed P1. However, it is also possible that this simulation disallows to sample the collapse of this partial P1 helix because it has the shortest time scale (640 ns). Unlike model 51, the stabilization of the P1 region in 3NPB without SAM can be supported by neighboring Watson-Crick base pairs and there is no competition from the AT helix.

4.3.5 Construction of the Energy Landscape from Multiple MD Simulations

Two different methods are employed here to visualize the free energy landscape based on multiple MD simulations:

First, the fraction of hydrogen bonds in Watson-Crick base pairs from the P1 helix and that from the AT helix are used as generalized coordinates. Figure 4.13 displays the free energy landscape (FEL) based on three different simulations on model 51—a simulation starting from model 51 only, a simulation from model 51 in complex with SAM and a restarted simulation of model 51 with SAM from the former simulation at frame 6615 with a complete P1 helix that has
the lowest RMSD to the counterpart in the crystal structure. The results suggest that model 51 locates at a branch point on the FEL—the formation of a stable AT helix (high $Q_{AT}$) is favored in the absence of SAM, while the presence of SAM allows model51 to navigate to other transient states and eventually leads to sample the conformation with a complete P1 (high $Q_{P1}$). However, the event of complete P1 helix formation is short-lived. Therefore, a restarted simulation from a snapshot with complete P1 is performed to evaluate the stability of this conformation. A 884 ns trajectory starting from frame 6615 of model 51+SAM simulation confirms the brief appearance of this conformation is not due to the instability of this particular conformation. The blue dashed arrows in Figure 4.13 is meant to demonstrate that this simulation only samples the bottom part of a deep energy funnel composed an ensemble with complete P1 helix, which should be the global free energy minimum in the presence of SAM. This feature is not reflected in this FEL visualization because the restarted trajectory has a shorter time scale compared to model 51+SAM. The other thing to note is that the transition from the conformation with 4 base pairs to complete P1 (6 base pairs) bypasses the conformational state with 5 base pairs, which suggests the helix formation does not follow a stepwise pathway but a cooperative pathway. But this should be confirmed with simulations with a shorter time interval (< 200 ps) to save the snapshots.

Additionally, PCA is employed to visualize the free energy landscape as proposed in reference 235. Again, mapping of model 51 on the FEL shows that this model resides at a branch point on the FEL. The global minimum is well featured here because two relative long trajectories (3NPB+SAM and 3NPB-SAM) are included in generating this FEL. This FEL representation also suggests that the most populated conformation in model 51+SAM—4 base pairs in the P1 helix occupies a local minimum with a small energy barrier to the global minima.
The fact that model 55+SAM has not sampled the complete P1 can also be visualized on the FEL—a high energy barrier between the energy funnel sampled in simulations on model 55 and the global energy minima. But it should be cautious that this may just due to insufficient sampling.

Figure 4.13 Energy landscape visualization for simulations on model51. Free energy landscape (FEL) view of simulations on model 51. These simulations include trajectory starting from model51+SAM complex, trajectory starting from model51 only and trajectory restarted from frame6615 (1.323 µs) from the first trajectory. The yellow star symbol indicates the position of the starting structure on the FEL. The dash arrows show some major transitions in temporal order represented in a color scale from black to light gray. The schematic transitions in the restarted trajectory from frame 6615 are shown in blue dash arrow. The fractions of hydrogen bonds in the P1 helix (Q_{P1}) and in the AT helix (Q_{AT}) are chosen as generalized reaction coordinates. The depth of the energy funnel is proportional to the population of conformations on the corresponding reaction coordinates. The local switching regions from some representative structures are shown in cartoon representation with P1 in blue and AT in red. SAM is displayed in sphere model.
Figure 4.14 Energy landscape visualization from PCA of the P1 helix. This figure is another representation of Figure 4.11b. The deep of the energy funnel is proportional to population of conformations at the corresponded reaction coordinates. Here the projections on the first two eigenvectors are used as the reaction coordinates. (Left) View of the FEL from the angle perpendicular to the plane of PC1 and Energy. The surface is plotted with a transparency factor alpha=0.5. (Right) Top view of the FEL.

4.4 Discussions

4.4.1 Transient State Modeling

The modeling of RNA tertiary structure with limited experimental data has come of age\textsuperscript{236}. The importance of RNA structure-function relationship motivates the development of theoretical models and implementations of computational tools for RNA tertiary structure modeling \textsuperscript{207,208,210,237-241}. One of these tools, MC-sym \textsuperscript{210}, has gained popularity due to the established pipeline for structural modeling and the ability to model relatively large RNA. This tool has been widely used to model RNA structures to fit SAXS data \textsuperscript{201,242}. In addition, a recent study utilized this tool to sample the tertiary contacts in different folding intermediates to explore the folding pathway of H\textdelta V ribozyme\textsuperscript{212}. One problem raised in this study is that the structural features sampled from MC-sym are lost during the energy minimization step. This problem can
be avoided by tweaking the parameter setups for energy minimization as suggested in a study using MC-sym to sample a wide range of tertiary structures for a tRNA to explore the constraints of different secondary structure topologies on RNA 3D structures\textsuperscript{226}.

In this study, the objective is to use MC-sym to generate 3D structures for some transient state models from secondary structure prediction and SAM binding experiments. Transient states are very difficult to characterize experimentally due to the short-lived, less populated features. However, transient states are very important for understanding some enzymatic reactions and folding of biomacromolecules. Therefore, computational approaches have become the major workhorse to tackle these problems. Using computational techniques to model a proposed transition state of Diels-Alder reaction, the researchers are able to generate theozymes and then engineer these catalytic sites into protein scaffolds to de novo generate an artificial enzyme\textsuperscript{243}.

The insights gains from transient state modeling using computational approaches will facilitate the experimental design to utilize some state-of-art experimental techniques\textsuperscript{244} to explore these less populated but important transient states.

4.4.2 Conformational Sampling

Although great advancement has been achieved to speedup MD simulations, a complete simulation of folding/unfolding for RNA of this size (~40 kDa) is still not possible. This is because the nucleation of the transient state takes up most of the folding time, while the propagation step can be hundreds to thousands of times faster than the overall folding rate [reference?]. Therefore, MC-sym is used to sample a discrete conformational space aimed to identify candidate transient state models with atomic details to bypass the most time-consuming part of the simulation. The results from PCA (Figure 4.11) demonstrate the ability of MC-sym to
complement the limitation on the conformational space sampling by the time scale in MD simulations.

MD simulations have become a very powerful tool to explore the motions of biomacromolecules at the atomic level. It is demonstrated in Chapter 2 that MD simulations can be used to identify some intrinsic instabilities in the riboswitch structure when the ligand is removed. It is admitted that high-resolution starting structures from X-ray or NMR are required in order to get reliable results from MD simulations. However, this problem can be alleviated with appropriate modeled structures. One previous study has showcased the capability of using 10-100 ns trajectories from MD simulations to perform the final structure refinement for protein structures from homology modeling\textsuperscript{245}. Additionally, a recent study has utilized MD simulations to evaluate the stability of some \textit{in silico} models for a membrane protein\textsuperscript{246}. Moreover, MD simulations have also been used to explore the conformational space for different \textit{in silico} mutants for an H\&V ribozyme to explore the role of several important residues in maintaining the active site geometry\textsuperscript{247}.

4.4.3 Advantages, Limitations and Problems in Current Framework

MC-sym has demonstrated the ability to sample a wide range of the discrete conformational space. However, this also raises the problem in choosing proper starting structures for MD simulations. Several analyses have been suggested in the MC-pipeline to evaluate structure models\textsuperscript{248,249}, and a recent study has used clustering to characterize these models\textsuperscript{226}. Since the structure modeling in this study is to be coupled with MD simulations, the same force field is used to evaluate models generated from the secondary structure constraints for a possible transient state. After that, the experimental results are used to prioritize the models
for the downstream MD simulations. The MD simulations part is the rate limiting step when the machine, Anton, specialized for speeding up MD simulation is not available.

Models from MC-sym indicate two possible penetration points for the 5’ strand of the P1 helix to occupy the switching regions—major groove of the AT helix and minor groove of the AT helix. Because the major groove of A-form helix is deep and narrow, it is difficult to sample a model with the 5’ strand placed in the major groove and still maintain good stacking among A109, U110 and A111 residues. Incorporation of fragments from a recently solved crystal structure of a triple helix233 may improve the modeling of this type of geometry.

One problem recognized in this study is that the FEL constructed here inherits the time scale limitation even though multiple MD trajectories have been used. The problem will be relieved when some conformational space is reevaluated with some restarted simulations, but this again will require highly intense computational resource. The other way to avoid this problem is to adapt weighted histogram analysis method 250 for multiple MD simulations.

4.4.4 Biological Implication

The study here suggests a possible folding intermediate state for yitJ SAM-I riboswitch. The simulation of model 51 in the absence of SAM maintains the AT helix structure. In addition, the simulation starting from model 51 with SAM navigates a pathway towards the formation of complete P1 helix, and restarted simulation origins at a conformation with complete P1 helix verify the stability of this conformation. The visualization of this transient state candidate and the transient ensemble of intermediate conformers from MD simulation advance the understanding of current mechanistic model for the function of SAM-I riboswitch. It is likely that the structural ensemble of yitJ SAM-I riboswitch includes some other hybrid conformations—partial P1 and
partial AT helix, for example, 3P1/10AT, 4P1/9AT, 5P1/8AT etc. (Boyapati et al, manuscript in preparation). Some of these species, such as the 6P1 conformation, would displace a sufficient segment of the 3’ strand of the AT to nucleate the downstream Terminator sequence, which explains the observation of basal transcriptional termination level $^{44,87}$. On the other hand, conformations with less base pairs in the P1 helix may not be stable enough to prevent the formation of the AT helix. However, this situation is shifted towards formation of the P1 helix in the presence of SAM. The simulation of model 51 in this study shows that the stabilization of the partial P1 helix by SAM anchor the 5’ strand of the P1 helix in an orientation that enables this single strand region to compete over the AT helix. This model is consistent with an NMR study on a small RNA system showing that the stabilization of a pre-formed helical region by a tetraloop increases the exchange rate between two different hairpin folds $^{251}$. 
CHAPTER 5 STRUCTURAL MODELING OF A SAM-I RIBOSWITCH IN THE “ON” STATE FROM NMR AND SAXS
5.1 Introduction to Solution State Biophysical Methods

The spatial folding of an RNA sequence is essential for its biological function\textsuperscript{252}. Structure determination by X-ray crystallography has provided a plethora of structures for riboswitches facilitating our understanding of the specific recognition between RNA and ligands\textsuperscript{35,71-79,81,82,84,87}. However, these conformations captured in crystal structures with the expression platform truncated may only represent one of the functional outcomes of riboswitches, but not the dynamics of the complete riboswitches. It is starting to be recognized that the functions of RNA are exerted through their unique dynamics\textsuperscript{253}. For example, transient interactions during folding have been shown to be crucial for the proper function of RNA\textsuperscript{197}. A recent single molecule study has suggested that an ensemble of multiple stable conformations is a more realistic description of the native state\textsuperscript{254}. Nevertheless, the conformational flexibility of RNAs create technical challenges for understand their functions. In Chapter2-4, computational tools have been employed to probe the RNA dynamics and the relevance of these transient events in the biological functions. In this chapter, several solution state biophysical methods have been utilized to study the mechanism of the SAM-I riboswitch and to explore the coupling with computational techniques.

5.1.1 Thermal Denature Method

Thermodynamics methods can be used to examine the folding of RNA. This type of approach has been used to gain insights on the structure of tRNA about five decades ago\textsuperscript{255}. Later on, melting studies have been widely employed to examine the structural stability of RNA under different pH\textsuperscript{256}, salt conditions\textsuperscript{257}, and study the thermodynamics of specific interactions between RNA and ligands and identify residues contributing to tertiary interactions in the RNA\textsuperscript{257}. In this study, one of these thermal methods—UV melting has been used to examine the
different melting behaviors of some RNA constructs representing the “ON” and “OFF” state of the SAM-I riboswitch and isolated structural elements within these conformers.

5.1.2 Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectroscopy is one of the most powerful experimental tools to study the dynamics of biomacromolecules\textsuperscript{258}. Recently, NMR spectroscopy, in combination with computational techniques, has manifested its power to characterize fleeting events that have functional relevance for optimization of enzyme activity\textsuperscript{259}. To achieve these goals, preliminary resonance assignment should be performed. NMR experiments for initial peak assignment are now introduced briefly.

The chemical shifts of imino protons locate in a so called RNA “fingerprint” region, of the NMR spectrum (15-9 ppm) that is well separated from other proton resonances. In some cases, imino proton signals may not be observed because protons exchange rapidly with water (millisecond time scale or faster). However, when the imino proton is involved in a base pairing, the exchange rate with water is significantly decreased, and thus the resonance from imino proton can be observed. The imino proton of U in a Watson-Crick A-U base pair resonates at 13-15 ppm, and the G imino in a Watson-Crick G-C base pair resonates at 12-13.5 ppm. Imino protons involved in noncanonical base pairings usually appear in the 9-12 ppm region. For example, G imino in a G-U base pair resonates at 10-12 ppm, and U imino at 11-12 ppm. Therefore, coarse grain structural information, such as the number of A-U or G-C base pairs, can be quickly gained from an imino proton spectrum. Additionally, stability of RNA secondary structure formation at base pair resolution can be obtained with temperature course and ligand titration 1D imino spectra respectively.
Further local structural information at atomic level—the distance between protons—can be extracted from the NMR NOESY experiment. This NMR experiment is based on the nuclear Overhauser effect (NOE)\textsuperscript{260}. This phenomenon shows that magnetization can be transferred between one dipole and another and the rate of magnetization transfer is inversely proportional to the distance between two protons to the 6\textsuperscript{th} power\textsuperscript{261}. Therefore, only local geometrical information, distance relationships for protons within 5-6 Å of each other, can be obtained from NOESY. The distance between imino protons from neighboring base pairs in the helical region is about 4 Å so that cross-peaks due to NOEs between these protons in nearest base pairs can be identified in the imino-imino proton region on the NOESY spectrum. Thus, initial peak assignments (mostly imino protons) and the RNA secondary structure can be obtained from the imino proton spectrum. In addition to local structural information from NOEs, another phenomenon—the dipolar coupling can be utilized to characterize global structure. The residual dipolar couplings (RDCs) arising from partially aligned samples in solution without distorting the spectrum can provide information for long-range distance, orientation and dynamics of biomacromolecule\textsuperscript{262,263}.

5.1.3 Small Angle X-ray Scattering (SAXS)

SAXS is another biophysical method that can be used to obtain structural information for biomacromolecules in solution. This experimental tool has become popular in the past several years mainly due to the availability of more SAXS beamlines at synchrotron light sources\textsuperscript{264}. SAXS is traditionally recognized as a tool to characterize macromolecules in solution state to gain size information, such as molecular weight, radius of gyration, and maximum intramolecular distance\textsuperscript{265}. Advancement of computational algorithms and their implementation has allowed using SAXS to generate global structural models of macromolecules\textsuperscript{266}. Moreover,
improvement of instrumentation enables the study of RNA folding within the millisecond timescale\(^\text{267}\). Also this tool has been shown to be a good complementary to NMR experiments, which can provide local distance information at atomic resolution, to model the global folding of RNAs\(^\text{268}\).

### 5.2 Experimental Procedures

#### 5.2.1 RNA Sample Preparation

DNA templates for in vitro transcription were generated by overlap PCR using overlapping oligonucleotides (IDTDNA). Primer with two consecutive 2’-O-methyl nucleotides (IDTDNA) at the 5’ end was used for the template strand\(^\text{269}\). RNA samples were then prepared from in vitro transcription using T7 polymerase\(^\text{270}\). Full length RNAs were separated from aborted transcripts using denaturing PAGE in 1 × TBE buffer. The full length product was observed under UV-shadowing and cut off the gel. RNAs were electroeluted from excised gel slices in 0.5 × TBE buffer, followed by ethanol precipitation. The RNA pellet was dissolved in MilliQ water and the samples were dialyzed extensively with series of dialysis buffers (1M, 0.5M, 0.2M and 0.05M NaCl, 10 mM sodium phosphate, 0.2 mM EDTA, pH=5.7). The RNAs were concentrated and desalted using Amicon 10K (Millipore). Final RNA stocks were stored at -20 °C. \(^{13}\text{C}/^{15}\text{N}-\text{labeled RNAs were prepared using the above protocol with NTPs purchased from Cambridge Isotope Laboratories Inc.}.

#### 5.2.2 NMR Experiments

The RNA stock was heat at 90 °C for 3 minutes and snap cooled on the ice for 5 min. Then NMR samples were prepared by dialyzing the RNA stock into NMR buffers using Amicon (Millipore). NMR experiments were performed on Varian 700 MHz spectrometer equipped
with cryo-probe, or Inova 500 MHz. Experiments involving exchangeable protons were generally collected in 90% H\textsubscript{2}O/10% D\textsubscript{2}O at specified temperature. Samples for non-exchangeable resonances were lyophilized at least twice in D\textsubscript{2}O. The experiments were recorded in 99.99% D\textsubscript{2}O at 15 °C. 1D spectra were processed in MestReNova LITE and other spectra were processed using NMRPipe\textsuperscript{271}.

5.2.3 UV Melting Experiments

UV melting experiments were performed in a Gilford Response II spectrophotometer equipped with a six-position thermal control cuvette holder. RNA samples are diluted to the buffer specified in cuvettes with stopper and incubated for 15 min at 70 °C in the spectrophotometer. The start temperature of heating experiment was 5 °C and the end temperature was 95 °C with a temperature step of 0.5 °C/min. The cooling experiment was performed immediately after the heating experiment. The following two criteria are used as quality control for data set: (1) no hysteresis between heating curve and cooling curve, (2) the absorbance difference at 5 °C is less than 2%.

5.2.4 Gel Filtration Column/Size Exclusion Chromatography

Gel filtration was performed with a Superose 12 10/300 GL column (GE healthcare) on the AKTA FPLC system equipped with a single wavelength UV-VIS detector (280 nm). The RNAs were purified in 20 mM MOPS (pH=6.5), 50 mM KCl, 7.6 mM MgCl\textsubscript{2} as the running buffer as recommend in reference \textsuperscript{272}. An additional run in the presence of ligand was performed with 100 µM SAM in the same running buffer. 1 mL of the running buffer was saved at the end of each run to be used as buffer for corresponding RNA sample during SAXS data collection.
5.2.5 Small Angle X-ray Scattering (SAXS) Experiment

SAXS data were collected at the X9/X21 beamline from the National Synchrotron Light Source (NSLS) at Brookhaven National Laboratory. Data were collected using a robotic arm to transfer the samples from a 96-well plate. Only one freeze-thaw cycle was allowed to ensure the quality of data collection. The RNA samples were spun at 14,000 x g for 10 min before data collection to remove air bubble in the sample. The samples were flowed through a capillary kept at 14 °C with a series of 1 s exposures to reduce the radiation damage. Data were collected in the following order—the empty capillary, the buffer and the RNA samples. Three replicated 30 second exposures were taken for each buffer and RNA sample. Raw images were examined manually to avoid the noise due to air bubbles. SAXS profiles were then determined with qualified images by subtracting the scattering due to the buffer and the empty capillary from the RNA scattering. Raw data were processed using pyXS (http://x9.nsls.bnl.gov/software/pyXS.htm) developed at NSLS. Data collection on a standard lysozyme sample was performed for every data collection trip for calibration to control for variation due to the energy fluctuation at the beamline.

5.2.6 RNA Tertiary Structure Modeling

The MC-sym pipeline is used to sample RNA conformational space based on known RNA structures\textsuperscript{210,242}. Models for this RNA construct are built based on the assumption that the helical part of RNA from P2 to P4 has the same fold as the crystal structure of the aptamer domain (PDB ID: 2GIS). The conformations of junction regions (shown in red) and AT helix (shown in orange) parts are sampled using MC-sym (Figure 5.11a). The conformational sampling is done without constraining the coaxially stack between the P4 and the AT helix.
Energy minimizations (max step is 2000 or frms < 1.0) were performed on the atomic structures of the models generated from MC-sym runs using Nucleic Acid Builder (NAB)\textsuperscript{223}. AMBER99bsc0 force field\textsuperscript{224} and Generalized Born model \textsuperscript{225} with an inverse Dybye-Hückel length of 0.19 Å\textsuperscript{-1} \textsuperscript{226} were used in the energy minimization procedure. 1000 models were generated in this step. This energy minimization protocol is mainly to rebuild the chain connectivity for models generated from MC-sym without introducing the sampling effect of the force field.

5.3 Preliminary Results and Discussions

5.3.1 Conformation Stability of Isolated AT helix Evaluated by UV Melting and NMR Experiment

To understand the role of AT helix in the SAM-I riboswitch, several isolated AT helices have been generated for UV melting and NMR experiment. Figure 5.1 presents the melting profile of an isolated AT_U2 (AT stands for anti-terminator helix, and two Us at the 3’ end is annotated as U2) helix. This RNA has a classical two state melting behaviors with a T\textsubscript{m} of 66.3 °C in 10 mM KPhosphate, 10 mM KCl, 0.02 mM EDTA, pH=6.0 buffer. The enthalpy calculated from this melting curve is -73.10 kcal/mol, which is very close to the enthalpy predicted using the nearest neighbor model with Turner’s thermodynamics parameters\textsuperscript{136}. Interestingly, another RNA, isolated AT_U4 helix (with two additional Us at the 3’ end), demonstrates a slightly different melting behavior compared to isolated AT_U2. As shown in Figure 5.2, the melting curve can be fitted into a model with two sequential transitions (see methods). The fitting yields a low temperature transition at \(\sim37.7\) °C and a high temperature transition at \(\sim60.2\) °C. Enthalpies extracted from the melting curve for these two transitions, -31.3 kcal/mol and -79.2 kcal/mol, are close to the predicted enthalpies for the two helical
Figure 5.1 UV melting of isolated AT_U2 helix. (a) Melting profile of isolated AT_U2 helix at 260 nm in buffer condition—10 mM KPhos, 10 mM KCl, 0.02 mM EDTA, pH=6.0 (similar condition as the buffer used in the NMR experiment without Mg²⁺). The dashed lines represent the upper (red) and lower (blue) baselines that are used for normalization. (b) Fraction (F) of folded state versus Temperature. Here \( F = \frac{A_U(T) - A_L(T)}{A_U(T) - A_L(T)} \), where \( A_U(T) \) and \( A_L(T) \) are linear equations for the upper and lower lines as shown in (a) respectively, and \( A(T) \) is the absorbance of the RNA sample under different temperatures. This analysis yields a \( T_m \) of 66.3 °C. (c) Sequence in secondary structure presentation of isolated AT_U2 helix. (d) Thermodynamic parameters determined by fitting \( \ln K \) vs. \( 1/T \). Here \( K = \frac{F}{1-F} \).

The slope of the linear fitting is \( -\frac{\Delta H^0}{R} \) and the intercept is \( \frac{\Delta S^0}{R} \), \( R \) is gas constant (1.985877534 cal·K⁻¹·mol⁻¹).
Figure 5.2 UV melting of isolated AT_U4 (a) Melting profile of isolated AT_U4 helix at 260 nm in 10 mM KPhos, 10 mM KCl, 0.02 mM EDTA, pH=6.0. The dashed lines represent the upper (red) and lower (blue) baselines that are used for normalization. (b) First derivative of the melting profile in (a). Black dashed line is the fitting assuming two sequential transitions (see details in method section). Individual transitions are shown in red dashed line and blue dashed line respectively. The following thermodynamics parameters are obtained from the fitting—\( \Delta H_1 = -31.3 \text{ kcal/mol} \), \( T_{m1} = 37.7 \degree C \) (transition 1); \( \Delta H_2 = -79.2 \text{ kcal/mol} \), \( T_{m2} = 60.2 \degree C \) (transition 2). (c) Two possible secondary structure models for isolated AT_U4. The enthalpies for these helical regions highlighted in the rectangle boxes are calculated using nearest neighbor model with Turner’s parameters—-84.7 kcal/mol for helix A, -30.9 kcal/mol for helix B, -59.2 kcal/mol for helix C.

regions in model 1, -30.9 kcal/mol and -84.7 kcal/mol (Figure 5.2c). The results suggest that this RNA is likely to adopt the secondary structure as model 1 in which helix B melts out at a low temperature, and helix A at a high temperature.

Additionally, NMR experiments were used to characterize these isolated AT helices. Imino protons are assigned by sequentially walking though the imino signals from base pairs in the helical region (Figure 5.3). Overall, these spectra display a very similar profile for base pairs presented in all the RNA constructs (Figure 5.4). The differences in the spectra among these RNA constructs lay are due to the following sequence differences. The construct with a wild type loop does not display the distinct imino peak from the G in UUCG tetraloop (~9.7 ppm), and the
Figure 5.3 Imino-imino proton region in NOESY NMR spectrum of different isolated AT helices and assignment of some imino protons. NOESY (mixing time = 300 ms) spectrum of different isolated AT helices in 10%/90% D2O/H2O at specified temperature. The lines connecting peaks in the spectrum represent the sequential NOE pathway for imino protons in helical regions. The imino protons are assigned to the residues in the 1D spectrum on the top and the sequence of each construct is shown on the right. (a) Isolated AT helix at 7 °C in 20 mM NaPhosphate, pH=6.0; (b) isolated WT AT helix at 7 °C in 20 mM NaPhosphate, pH=6.0; (c) isolated AT_U4 helix at 5 °C in 5 mM KPhosphate, 10 mM KCl, 0.02 mM EDTA, pH=6.7; (d) isolated AT U2 RNA at 15 °C in 10 mM NaPhosphate, 10 mM NaCl, 0.2 mM EDTA, pH=6.0. Constructs in (a) and (b) are artificially stabilized with GC pairs at either end in (a) or at one terminus in (b). Constructs (a), (c) and (d) are with a UUCG tetraloop.
Figure 5.4 Comparison of the 1D imino proton region of different AT helix constructs. The spectra are listed from top to bottom in the following order: AT_helix with wild type loop (pH=5.8), AT helix (pH=6.0), AT_U2 (pH=6.0) and AT_U4 (pH=6.7).

Figure 5.5 NMR temperature course experiment for isolated AT helix with wild type loop and isolated AT_U2. (Left) Isolated AT helix with the WT loop (sequence as shown in Figure 5.3b. (Right) Isolated AT_U2 (sequence as shown in Figure 5.3d).
imino signal from U90 is not observed because of the instability of the wild type loop. Two closing GC base pairs at the end of the stem (AT helix and WT AT helix) appear to stabilize the region around the UC mismatch for those two constructs. NMR experiments for isolated AT_U2 and AT_U4 constructs suggest that these RNA constructs are more likely to adopt the secondary structure model 1, with a UC mismatch to separate two helical segments instead of a GU bulge in model2, as demonstrated in Figure 5.2c. Utilizing the base pair information from NOESY experiments, NMR temperature course experiments show that base pairs around the UC mismatch region melt out first. It is likely that the vulnerability of the AT helix, the UC mismatch region, may be the invasion point for the 5’ P1 strand, and thus this could represent a switch point in the functional decision step of the SAM-I riboswitch (as discussed in Chapter 3).

5.3.2 Structural Information on the Engineered “ON” State Conformers

Several engineered “ON” state conformers, with the 5’ strand of the P1 helix truncated to constrain the RNA into formation of conformers representing a possible “ON” state, were submitted for different solution state biophysical. Here the annotation for these ON state conformers uses a unique string of nucleotides at the 5’ end and the number of Us at the 3’ end to distinguish different RNA construct. For example, GAGG_AT_U2 means the RNA construct starts with GAGG at the 5’ end with an AT helix in the sequence, and the 3’ ends with 2 Us.

First, the melting profiles for an “ON” conformer—GAGG_AT_U2 and the “OFF” conformer—AAT were compared (Figure 5.6). Under low monovalent ionic concentration (10 mM and 100 mM KCl), these two RNAs present two distinct melting profiles—AAT has a dominant transition at low temperature range, while the hyperchromicity of GGAG_AT_U2 has more contribution from the transition at high temperature. In the presence of
Figure 5.6 Comparison of melting profiles for an "ON" conformer and the "OFF" conformer. (a) Melting curves and first derivatives of the curves for GGU_AT_U2 (ON conformer) and AAT (OFF conformer) in 10 mM KPhosphate, 10 mM KCl, 0.1 mM EDTA, pH=7.0. (b) Same as (a) but with 2mM Mg$^{2+}$ in the buffer. (c) Schematic of the two RNA constructs used for comparison.

Figure 5.7 Melting profiles for GGU_AT_U2 RNA construct under different ionic conditions. (a) Fitting of the normalized melting curves under different ionic conditions at pH 6.0 (10 mM NaPhosphate). The raw data is shown in open circles. (b) First derivative of the fitted melting curves. (c) T_m versus salt concentration for two assumed sequential transitions from melting curve fitting. (d) Ratio of enthalpy associated with the 2nd transition to that with the 1st transition under different salt conditions. (e) Sequence of GGU_AT_U2 RNA in secondary structure representation.
Mg$^{2+}$, these two RNAs demonstrate relatively similar melting profiles. The results suggest that these conformers represent two different states of the SAM-I riboswitch having different ionic dependence in their folding behaviors. Mg$^{2+}$ may be more important for the stability of AAT than for GAGG_AT_U2, which includes an AT helix that has been shown to melt out at high temperature even without Mg$^{2+}$.

Figure 5.7a-b show the UV melting profiles of GGU_AT_U2 RNA. The results clearly display multiple transitions occurring during the unfolding of RNA. Here the data is only fit to two sequential transitions attempting to dissect different effects of ionic conditions on individual transition. The results show that melting temperatures for these two transitions have different salt dependence—transition 1 (high slope) is more sensitive to the concentration of monovalent ions than transition 2 (low slope). However, the ratio of the van’t Hoff enthalpies for these two transitions changes as the salt concentration increase. Previous study has shown that unfolding of duplex RNA enthalpies are not strongly salt dependent$^{273}$. This suggests that one of the two transitions is not a complete two-state transition at lower salt concentrations. It is also possible this is due to the inadequate fitting of only two transitions to the unfolding of a such large, complicated RNA compared to small, hairpin RNA. Therefore, more advanced models describing the unfolding pathway of large RNAs should be used to fit the melting curve within the sensitivity of the instrument or melting behaviors under different wavelengths could be incorporated to gain more information about large RNAs. These solutions may make UV melting experiments more informative for studying large RNAs$^{274}$.

Figure 5.8 displays some preliminary results from NMR experiments on one of the large RNA constructs—GGAA_AT_U2 representing a possible ON state. With the imino proton assignments on the isolated AT helix, most corresponding imino protons can be assigned in the
Figure 5.8 NMR experiments on GGAA_AT_U2 RNA construct. (a) Imino to imino proton region of 1 100 ms NOESY spectrum of GGAA_AT_U2 RNA in 10 mM NaPhosphate, 10 mM NaCl, 0.2 mM EDTA, pH=6.0 at 15 °C. Initial assignment of the imino proton is shown in the 1D imino spectrum on the top. (b) $^{15}$N Trosy HSQC of $^{13}$C, $^{15}$N labeled GGAA_AT_U2 RNA under the same condition as (a). (c) Superposition of imino to aromatic/amino proton region from the NOESY experiments of GGAA_AT_U2 (peaks in red) (same as (a)) and the 300 ms NOESY spectrum of isolated_AT_U2 (peaks in blue). Cross peaks corresponding to AU base pairs are highlighted in a purple box, a subset of those associated with GC pairs is in a cyan box, and peaks from GU pairs in a green box. (d) Sequence of GGAA_AT_U2 RNA in secondary structure representation. Residues with imino proton assigned are highlighted in red.

large RNA construct with some additional signals from other helices also being assigned (Figure 5.8a). The assignments are further confirmed with $^{15}$N HSQC (Figure 5.8b), which yields the direct information about the connection between the proton and the cyclic nitrogen. Additionally,
Figure 5.9 Comparison of 1D imino proton spectrum for isolated AT_U2, GGAA_AT_U2 and isolated P4 helix at 15 °C. Initial imino proton assignment is based on the information from GGAA_AT_U2 construct. Peaks in the isolated P4 helix that line up with GGAA_AT_U2 are assigned in black. Other peaks can be assigned then based on the secondary structure model of the P4 helix—peaks having no conflicts with the assignment in GGAA_AT_U2 are shown in green, one peak, U74, that is inconsistent with the assignment in GGAA_AT_U2 is shown in red.

The superposition of the imino-aromatic/amino region from NOESY experiments between the large RNA and the isolated AT helix demonstrates that peaks from these two constructs overlap very well (Figure 5.8c). This suggests that the local geometry of the AT helix is the same in both constructs. The information from these NMR experiments appears to be consistent with the expected secondary structure model as shown in Figure 5.8d—with majority of the imino protons from the P4 and AT helix assigned, an expected over-crowded region of peaks from imino proton of GC base pair due to the highly GC rich P2 and P3 helix and some unassigned peaks in 10-11 ppm region for imino protons in noncanonical base pairs possibly from GA base pairs in the P2 helix. To further assist the peak assignment, an isolated P4 helix has been generated from NMR experiment (currently only 1D spectrum available). As demonstrated in Figure 5.9, most peaks in the imino proton region of isolate P4 helix do not have conflicts with the assignment in the large construct, except U74. Additionally, the broadening of imino protons from the P4 helix...
and other helices in the large RNA construct is an indication of an intermediate slow exchange in these regions.

SAXS is also employed here to characterize the global folding of these ON state conformers. Figure 5.10 shows the scattering profile of GAGG_AT_U2 RNA. The Kratky plot of SAXS data is a way to monitor the degree of compactness—bell-shaped curve for globular particles and plateau at high $q$-values for Gaussian chains. A Kratky plot for the sample in the absence of SAM is more parallel to the baseline at high $q$ region than that in the presence of Mg$^{2+}$, suggesting that the RNA has a higher degree of “unfoldedness” without Mg$^{2+}$. However, complete “foldedness”, characterized with a sharp peak at the low $q$ values, is still not observed with 2 mM Mg$^2$. A Guinier plot is used to extract size information about this RNA construct (Figure 5.11). The result shows that this “ON” state conformer has a significant compaction in the presence of 2 mM Mg$^{2+}$—$R_g$ decreases from 37.1 Å to 27.7 Å. To gain further structural information, we used an ab initio density reconstruction algorithm implemented in DAMMIN$^{275}$ to generate low resolution structural models from SAXS data for GAGG_AT_U2 under two different buffer conditions. These molecular envelopes represent the averaged electron density of the ensemble$^{276}$. This RNA adopts an elongated shape without Mg$^{2+}$, while it becomes more compact and has a “Y” shape conformation (Figure 5.11b).

A further step is taken to determine the RNA structure using SAXS data. This is now plausible due to the advancement in the field of RNA tertiary structure prediction with secondary structure constraints and the development of a theoretical framework to incorporate the effect of first shell solvent into calculating the scattering profile. Two strategies are employed here. One is direct fitting by predicting the scattering profile for
Figure 5.10 Scattering profile of GAGG_AT_U2 RNA construct under two different buffer conditions. (Left) Raw scattering profile of GAGG_AT_U2 RNA in 10 mM KPhosphate, 10 mM KCl, 0.02 mM EDTA, pH=6.0 without Mg2+ (black) and with 2 mM Mg2+ (red). (Right) Kratky plot of the SAXS data shown in the left figure.

an atomic structure, then comparing with the scattering profile to the experimental one. The other approach is to fit the atomic structures to the calculated coarse grain density map as shown above geometrically. Figure 5.12b shows the distribution of chi-square for MC-sym models using the direct fitting. For current tertiary structure modeling with the constraint from the secondary structure model in Figure 5.12a, the 1st rank model stands out from other models. Comparing the experimental and theoretical scatter profile, the slight discrepancy in the low q region indicates that the actual RNA is a litter more bulky than the 1st rank model even though the first shell water has been taken into consideration (Figure 5.12c). This may require an improved model for the water around RNA molecules. An interesting structure feature presented in the 1st rank model is that the co-axial stacking of the P2a stem, pseudoknot (PK) and the P4 helix (Figure 5.12d). The geometry around the binding pocket in the crystal structure, J3/4 and P3 helix as
Figure 5.11 Coarse grain model of an “ON” state conformer—GAGG_AT_U2 from SAXS experiment. (a) Guinier plot for the SAXS data shown in Figure 5.9. The data points in the region that $q \times R_g < 1.3$ are used to calculate the radius of gyration. (b) Coarse grain structural models for the GAGG_AT_U2 RNA construct under two different buffer conditions—no Mg$^{2+}$ (red) and 2 mM Mg$^{2+}$ (blue). The smallest unit of the grid shown in the background is 10 Å, the diameter of an A-form RNA helix is about 20 Å. The molecular envelopes presented here are generated by convoluting the bead models with a Gaussian kernel using Situs. 

---

108
Figure 5.12 Modeling atomic structure using SAXS data in the presence of Mg$^{2+}$. (a) Sequence of GGAG_AT_U2 RNA in secondary structure representation. This secondary structure is the constraint for tertiary structure modeling using MC-sym$^{210}$. (b) Cumulative model plot as a function of $\chi^2$. The $q$-range used in the calculation is 0.016-0.29 Å$^{-1}$. The data points are shown in blue open circle. (c) Comparison of the experiment SAXS data (black open circle) and Fast-SAXS-RNA$^{242}$ calculated (red) SAXS profile for the 1$^{st}$ rank model. (d) Stereo view of the 1$^{st}$ rank model in cartoon representation. The color scheme is the same as the labeling in (a). The bar is to show the co-axial stacking among the P4 helix, PK and the P2b helix.
Figure 5.13 Top 8 rank structure models from indirect fitting to coarse grain electron density. The coarse grain electron density calculated using the SAXS data in the presence of Mg\(^{2+}\) is shown as surface with transparent blue color. The backbone of the atomic models is shown as tube and the nucleotides are shown in stick representation. The following color scheme is used to distinguish different components in the structure—P2 helix in green, P3 helix in orange, J3/4 in magenta, P4 helix in blue and AT helix in red.

presented in this model, is collapsed, and thus does not allow SAM binding. Figure 5.13 displays the top 8 rank models for the alternative strategy. The consensus message from these models is that the AT helix cannot adopt the orientation relative to the P3 helix that the P1 helix does in the crystal structure to sandwich the ligand—SAM in between. More secondary structure constraints can be utilized to model the tertiary structure to improve the fitting of models to the SAXS data\(^{242}\). Moreover, some other experimental techniques with single nucleotide resolution information, such as hydroxyl radical probing\(^{278}\), can be applied to assist model construction or selection further.
5.3.3 Observation of SAM Effects on Full Length SAM-I Riboswitch Constructs

NMR has been shown to be a powerful tool for studying conformational dynamics. Here a large RNA construct—WT AAT_AT, including the full length sequence that can either form a P1 helix or AT helix, is generated for NMR study to find some characteristic peaks for dynamics study. Meanwhile, an isolated P1 helix and WT AAT RNA are also submitted for NMR experiment to assist the identification of signature signals from different conformations. As demonstrated above, relatively small isolated

Figure 5.14 NMR experiments on isolated P1 helix and WT AAT RNA. (a) Imino to imino proton region of a 250 ms H2O NOESY of P1 helix at 7 °C. Buffer: 20 mM NaPhosphate, pH=6.0. The sequence of the isolated P1 helix is shown at the right with a red rectangle box highlighting the local environment same as the P1 helix is the large RNA construct. Residues with imino proton assigned are shown in colors. (b) Superposition of the 1D imino proton of WT AAT and isolated P1 helix. The sequence of WT AAT is shown at the right in secondary structure representation. The red arrows highlight the imino proton signals from the only GU base pair in the P1 helix.
Figure 5.15 1D imino spectrum of different RNA constructs. The 1D imino proton spectrum of various *T. tencongensis metF* SAM-I riboswitch RNA constructs in 10 mM potassium phosphate, 10 mM potassium chloride and 0.01 mM EDTA (pH 6.0) in 90% H$_2$O/10% D$_2$O at 35°C under three different conditions—no Mg$^{2+}$, in the presence of 2 mM Mg$^{2+}$ and with 5:1 ratio [SAM]/[RNA]. (a) SAM-I riboswitch RNA construct containing “strand switching” elements capable of forming P1 or AT helix (b) Truncated “aptamer” sequence capable of forming only P1-helix containing conformers. (c) Construct truncated at the 5’ end to prevent P1 helix conformation. (d) Isolated AT helix construct. Addition of SAM reduces intensity of signal associated with GU base pair within AT helix, while increasing intensity of another signal previously observed in truncated aptamer. Note that whereas constructs used in parts a and b contain wild type sequences, RNAs used in panels c and d contain a UUCG tetraloop substitution in the AT helix, giving rise to a signal at 9.7 ppm. The signature signals from the AT helix and the P1 helix are highlighted in transparent yellow and pink rectangle box respectively. The positions of these two GU base pairs are also highlighted in the schematic on the right.
helix constructs can quickly yield imino proton assignment information (Figure 5.14a). With the aid of the isolated P1 helix, the signal from the only GU base pair in the P1 helix can be located in the spectrum of large RNA constructs (Figure 5.14). Next, NMR experiments are applied to detect the conformational change upon the addition of SAM.

Figure 5.15 presents the “fingerprint” region of different RNA constructs under three different conditions. The WT AAT, which can only form the aptamer domain, displays no significant change when SAM is added. For the WT AAT_AT_U2 construct that can form either the AT or the P1 helix in a competition, a number of changes are observed in the spectrum upon addition of SAM. In particular, the signals from the GU base pair from the AT helix (Figure 5.3) and that from the P1 helix demonstrate a direct evidence of the equilibrium shift. Upon the addition of SAM, the signals from the AT helix GU base pair are reduced, while these from the P1 and some other signals line up with the WT AAT spectrum increase. These results suggest that the AT helix form is dominant in the absence of SAM, and the equilibrium is shifted toward formation of the P1 helix when SAM is present.

The direct evidence of a conformation equilibrium shift from NMR experiments helps the interpretation of gel filtration running profiles for RNA constructs with different 3’ end truncation point (Figure 5.16c). The running profile for WT AAT is consistent with the previous observation in the literature\(^{272}\). Based on the information from NMR experiment, the fast running peak of AAT_AT_U2 with increased population in the presence of SAM is interpreted as the “OFF” state/ligand bound conformer, and the 2\(^{nd}\) peak may represent the “ON” state conformers. With the assumption that AAT_AT2 should have an analogous elution profile as to that of AAT_AT_U2, the results show that the conformational equilibrium is shifted towards the “ON” state and becomes less sensitive in responding to the presence of SAM as the sequence extends at
the 3’ end. These observations are also consistent with the binding data from equilibrium dialysis presented in Chapter 3. The eluted fractions from the center area of individual peak were collected for SAXS experiment. The preliminary parameters extracted from SAXS scattering profiles are listed in Table 5.1.

Figure 5.16 SAXS scattering profiles and gel filtration column profiles for WT sequence with different 3’ end truncation points. (a) Guinier plots of SAXS data for two different RNA constructs in the absence and in the presence of SAM. (b) Kratky plots to examine the “foldedness” of the RNA samples. (c) Running profiles for three different RNA constructs
Table 5.1 $R_g$ and $D_{\text{max}}$ calculated based on the Guiner region of the 1D scattering profiles.

<table>
<thead>
<tr>
<th>Sample</th>
<th>$R_g$ (Å)</th>
<th>$D_{\text{max}}$ (Å)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAT – SAM</td>
<td>53.0±0.5</td>
<td>218</td>
</tr>
<tr>
<td>AAT + SAM</td>
<td>28.3</td>
<td>120</td>
</tr>
<tr>
<td>AAT_AT_U2 – SAM</td>
<td>59.0±0.6</td>
<td>270</td>
</tr>
<tr>
<td>AAT_AT_U2 + SAM</td>
<td>70.9±1.4</td>
<td>282</td>
</tr>
</tbody>
</table>

These observations have the following implications: 1) The truncated aptamer model is not sufficient for studying the ligand induced conformation exchange; 2) At least two conformational families exist in slow exchange for WT AAT_AT_U2, as indicated by the existence of separate signals. 3) Alignment between spectra from different constructs and comparison of spectra in the presence and in the absence of SAM are consistent with a conformational change between “ON” and “OFF” state conformers. 4) The SAM-I riboswitch is not well described by an “all or none” model of conformational switching. Rather, the ligand perturbs a delicately balanced equilibrium, allowing for fine tuning of the level of gene expression to achieve the function as “dimmer switch”. This is the first study to show that the SAM-bound conformer is not 100% in the aptamer form (at least at this transcript length. It does bring up a paradox. If the measured $K_d$ is nM, then at this concentration and with excess SAM almost 100% of the RNA should be bound. Does that mean that SAM is binding to some AT-containing RNAs?
CHAPTER 6 CONCLUSIONS AND FUTURE WORK
The work presented here focuses on structure and mechanism of the SAM-I riboswitch that regulates gene expression at the transcriptional level. Both computational and experimental approaches are taken towards this objective.

In Chapter 2 and reference\textsuperscript{130}, our results from MD simulations on the crystal structures of the SAM-I riboswitch aptamer domain suggest that the positioning of the J1/2 junction upon SAM binding helps to overcome the entropic barrier to formation of the P1 helix. SAM binding appears to facilitate the coordination of a magnesium ion to phosphates within the J1/2 strand. Interestingly, a non-adjacent dinucleotide stack that is present in the crystal structure of \textit{B.subtilis} \textit{yitJ} SAM-I riboswitch emerges in our simulations. This observation suggests that this interaction may form transiently in the \textit{T.tengcongensis metF} SAM-I riboswitch RNA as well.

However, variation in the regulatory response to SAM observed in several wild-type SAM-I riboswitches cannot be solely explained by the binding affinity of the aptamer domain\textsuperscript{44}. In Chapter 3, we proposed to consider Boltzmann distributions of riboswitch conformers to decode the variations in sensitivity to cognate ligand and in the level of modulating gene expression. The results from Co-transcriptional Folding Simulations show more than one plausible switching points for the decision regarding folding and functional outcome for the SAM-I riboswitches. Several decoy interactions with the potential to sequester the J1/2 region are not highly conserved but appear in subsets of SAM-I riboswitches. We propose that these variable interactions are likely to be intrinsic factors for the fine tuning mechanism, adjusting each riboswitch for the unique gene expression requirements of downstream genes. One of these decoy interactions, involving sequestration of J1/2 and part of the 5' portion of the P1 helix, is
visible but previously unnoticed in the literature. Additionally, we have provided indirect experimental evidence for a second decoy interaction involving the formation of a P0 helix. Our BPP calculations suggest that SAM binding can stabilize P1 helix formation by contacting G11, thus blocking formation of the “misfolded” decoy structures involving G11 base pairing. These findings are also in line with the results from MD simulations on the aptamer in Chapter 2 and reference 130. Preliminary experimental results in Chapter 5 pave the way for further characterization of the dynamics of the SAM-I riboswitch. Significantly, peaks to directly monitor the competition between the P1 helix and the AT helix have been pinpointed in current NMR spectroscopy. Identifying a suitable model system for structural analysis by NMR is often a tedious process of trial and error, especially for a large, complex RNA. The customary means of optimization is by truncating dynamic segments as for X-ray crystallography. Yet in this instance we found that the most informative model system contains the elements required to mimic the switching mechanism, once spectra for the isolated fragments are available to aid the assignment process.

The impact of SAM on a strand switching competition between the 5’ P1 strand and the 3’ AT strand is explored with transient state modeling, experimental observation and MD simulations (Chapter 4). Our MD trajectories of unprecedented time-scale for RNA of this size provided atomic details for a potential folding pathway of the SAM-I riboswitch. A model explaining the mechanism of SAM as the effector utilizing a small contact surface with the RNA to initiate the strand switching of the SAM-I riboswitch emerges from this study (Figure 6.1). Interestingly, the topology describing the competition between the P1 helix and the AT helix resembles that of certain deoxyoligonucleotide devices used in DNA computing279. In the artificial DNA devices, the computing rate can be adjusted by varying the length of the toehold
Figure 6.1 Schematic of thermodynamics cycle model for the SAM-I riboswitch and an example of oligo devices used in DNA computing. (a) A proposed thermodynamics cycle describing the relationship between SAM binding and the folding of SAM-I riboswitch. The strand segment in the 5’P1 that has contacts with SAM is shown in blue. The folding of SAM-I riboswitch in the absence of SAM is an equilibrium between an intermediate hybrid state (H/HB) and the “OFF” state (O/OB), characterizing by the free energy $\Delta G_{\text{hybrid-AAT}}$ and $\Delta G_{\text{hybrid-AAT}}^{SAM}$ in the presence and absence of SAM respectively. The interactions of SAM with the hybrid and the “OFF” states are described by the free energy $\Delta G_{\text{hybrid-AAT}}^{SAM}$ and $\Delta G_{\text{hybrid-AAT}}^{SAM}$ respectively. (b) Schematics for oligo devices used in DNA computing from reference. The example showing here includes two species—input and gate:output. Solid lines are used to represent DNA strands, the direction of the arrow is from 5’ to 3’, and the colors indicate different DNA strands. In this example, S5 is the recognition domain, T is a short toehold domain and T* is the Watson-Crick complement of T. Varying the length of the toehold domain can be used to adjust the performance of DNA devices.
domain\textsuperscript{279}. In the \textit{yitJ} SAM-I riboswitch, SAM modulates the functional outcome by stabilizing the toehold-like domain in the P1 helix.

This work has probed into the less conserved region—the expression platform of the SAM-I riboswitches—to sort out the source of functional variability within the sequence context. This follows the trend of recent studies in this field on the contribution of peripheral elements around the ligand binding pocket understates the innovative nature of this work\textsuperscript{174,280}.

Nevertheless, several open issues need to resolved in future work. Experiments, such as the recent SHAPE-seq\textsuperscript{63}, that can quantitatively measure the contribution from individual decoys will help dissect and correlate with variations in the functional outcome. Further dynamics study with NMR spectroscopy will provide quantitative information about the equilibrium shift perturbed by SAM. Additionally, our study in Chapter 4 is the very first study that has utilized current force field for such a long time scale. Such long timescales will become more readily available with the advance of computer hardware. Our trajectories indicate that these timescales can offer insights into interesting biological processes involving the propagation steps within a strand exchange process Further evaluation of the performance of force field for a large complicated RNA system to eliminate the artifacts therefore is now imperative.
REFERENCES


APPENDIX A: GENOMIC CONTENT DOWNSTREAM OF SAM-I RIBOSWITCHES
VITA

Wei Huang was born in June 1985, in Yongchun, Quanzhou, Fujian, China. He received his Bachelor of Engineering from Xi’an Jiaotong University in Xi’an, Shannxi, China, in July of 2006. He joined Dr. Aboul-ela’s lab at Louisiana State University in Baton Rouge, Louisiana, in August 2007. Wei Huang will graduate with the degree of Doctor of Philosophy in December 2011. He will attend Case Western Reserve University in January 2012 where he will conduct post-doctoral research on the nuclear estrogen receptor (ER) in Dr. Yang’s lab.