Study of complex RNA function modulated by small molecules: the development of RNA directed small molecule library and probing the S-adenosyl methionine discrimination between on and off conformational states of the SAM-I riboswitch

Vamsi Krishna Boyapati
Louisiana State University and Agricultural and Mechanical College, vboyap1@tigers.lsu.edu

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

Recommended Citation
https://digitalcommons.lsu.edu/gradschool_dissertations/2172
STUDY OF COMPLEX RNA FUNCTION MODULATED BY SMALL MOLECULES: THE DEVELOPMENT OF RNA DIRECTED SMALL MOLECULE LIBRARY AND PROBING THE S-ADENOSYL METHIONINE DISCRIMINATION BETWEEN ON AND OFF CONFORMATIONAL STATES OF 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

Vamsi Krishna Boyapati
B.S, Osmania University, India, 2003
M.S, Acharya Nagarjuna University, 2005,
December 2011
DEDICATION

I would like to dedicate this work to my father Niranjan Prasad Boyapati, mother Satya Vara Lakshmi Boyapati and to my friend late Dr. Kiran Kumar Allam.
ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my advisor, Dr. Fareed Aboul-Ela, for investing in me as a student. I would like to thank him for being a very supportive mentor over the past several years. I would like to thank him for his scientific insight and guidance and for all that he has taught me. I would like to extend special thanks to my advisory committee Dr.Anne Grove, Dr.Marcia Newcomer, Dr.Graca Vicente for their continuous support and guidance all through my graduate studies.

I would like to thank Dr.Dale Teleaven and Dr.Thomas Weldegheorghis for their assistance with nuclear magnetic resonance spectroscopy automation and pulse program optimizations for the RNA directed small molecule library development. I would like to thank Dr.Jacqueline Stephens for providing materials and allowing me to use their lab for S-Adenosyl methionine binding experiments. I would also like to thank past and present members of Dr.Aboul-Ela’s lab for their help and cooperation. I would like to thank all my collaborators including Dr.Joohyun Kim and Dr.Shantenu Jha for their helpful discussions and assistance. I would like to specially thank co-graduate student Wei Huang for his discussions and the work we shared in the SAM-I riboswitch project. I would like to thank Dr.Kiran Kumar Allam for his assistance in analyzing fragment molecules and similarity search in the RNA directed small molecule library project. I would like to thank Dr.Chandrashekhar Reddy Komma for his support and discussions.
I would finally like to thank my parents and my sister for their continuous support and patience.
# TABLE OF CONTENTS

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

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

LIST OF TABLES .................................................................................................................................... vi

LIST OF FIGURES ............................................................................................................................... vii

ABSTRACT ........................................................................................................................................... ix

CHAPTER 1: INTRODUCTION ................................................................................................................. 1
 1.1 RNA as Potential Small Molecule Drug Target ........................................................................... 1
 1.2 RNA Structural Features and Their Contribution towards Small Molecule Binding ................. 3
 1.3 Molecular RNA Targets ............................................................................................................... 7
 1.4 Small Molecule RNA Interactions ............................................................................................ 7
 1.5 Riboswitches Are Classic RNA Systems to Understand Diverse Small Molecule-RNA Interactions .......................................................................................................................... 11

CHAPTER 2: DESIGN AND IMPLEMENTATION OF AN RIBONUCLEIC ACID (RNA) DIRECTED FRAGMENT LIBRARY ............................................................................................................. 32
 2.1 Introduction ...................................................................................................................................... 31
 2.2 Methods and Materials ................................................................................................................. 33
 2.3 Results ............................................................................................................................................ 37
 2.4 Conclusion ..................................................................................................................................... 45

CHAPTER 3: DETERMINING THE STRUCTURAL ELEMENTS REQUIRED FOR LIGAND DISCRIMINATION BETWEEN ON AND OFF STATE SAM-I RIBOSWITCH CONFORMATIONS. ................................................................................................................................. 48
 3.1 Introduction ...................................................................................................................................... 48
 3.2 Methods and Materials .................................................................................................................... 53
 3.3 Results ............................................................................................................................................ 62
 3.4 Discussion ..................................................................................................................................... 79
 3.5 Conclusion ..................................................................................................................................... 87

CHAPTER 4: SUMMARY AND FUTURE DIRECTIONS ........................................................................... 89

REFERENCES ........................................................................................................................................... 95

APPENDIX: COPYRIGHT PERMISSIONS FROM JOURNALS .............................................................. 110

VITA .......................................................................................................................................................... 119
LIST OF TABLES

1. The thermodynamics of the RNA secondary structures..........................4
2. Statistics of the clustering and fragment database.................................39
3. The fragment hits from initial screen and competition screen.................43
4. List of *yitJ* SAM-I riboswitch hybrid constructs and their primers..........55
5. Restriction reactions of the plasmid and insert for SAM-I riboswitch.......57
6. Ligation reactions for restricted plasmid and insert................................57
7. Reported $K_d$ values for different SAM-I riboswitch constructs...............70
LIST OF FIGURES

1. A form and B form helices.................................................................5
2. Schematics of the RNA secondary structure elements..........................6
3. A cartoon representation of the Gentamicin C1A and the A-site RNA......10
4. Schematic representation of riboswitch mechanisms...............................16
5. The distribution of riboswitches in pathogenic bacteria.........................18
6. Lysine analogs...................................................................................21
7. The ligands designed for targeting the Guanine riboswitch....................25
8. S-Adenosyl Methionine and its analogs.................................................27
9. RNA polymerase schematic...............................................................28
10. Example of descriptive properties.........................................................39
11. Distribution of number of fragments per cluster....................................40
12. The secondary structure of the A-site RNA..........................................41
13. Screening data..................................................................................42
14. Follow up WaterLOGSY and T_2 filter experiments............................44
15. RNA Imino spectrum with library hit fragment....................................45
16. The secondary structure of the T.tencongensis metF SAM-I riboswitch....50
17. Secondary structure and sequence of B.subtilis yitJ SAM-I riboswitch.....51
18. T.tencongensis On constructs............................................................52
19. B. subtilis yitJ SAM-I riboswitch AT and AAT hybrid constructs.............53
20. Schematic representation of cassette PCR strategy...............................55
21. 2% ethidium bromide stained agarose gel for PCR reactions................56
22. Hammerhead ribozyme integrated SAM-I riboswitch hybrid construct...60
23. WaterLOGSY experiment for SAM with *T. tencongensis* aptamer ..........63
24. Equilibrium dialysis data for *T. tencongensis* aptamer .........................64
25. T2 filter SAM binding experiment with the GGAG_AT_U4 construct ........65
26. WaterLOGSY experiment for GGAG_AT_U4 construct ..........................66
27. Equilibrium dialysis data for *T. tencongensis* On constructs .................66
28. BPP calculation for *B. subtilis* yitJ SAM-I riboswitch ..........................68
29. Scatchard analysis ..............................................................................71
30. *B. subtilis* yitJ SAM-I riboswitch mutation analysis ..............................72
31. *B. subtilis* yitJ SAM-I riboswitch equilibrium dialysis ...........................75
32. 10% denaturing acrylamide gel for inline probing analysis ........................76
33. Equilibrium dialysis data for SAM-I hybrids to test stability of P1 helix ....77
34. Superposition of models generated from MC-sym .....................................80
35. A secondary structure representation for 8P1_5AT ....................................91
36. A model for SAM-I riboswitch folding in the presence of SAM ...............92
ABSTRACT

RNA recently remained unexploited and is now drawing interest as a potential drug target. The methodology and available drug libraries for RNA targeting/screening are in rudimentary stages. The interactions made by ligands with RNA can be explored for RNA based drug development. The dissertation is composed of 4 chapters. The first chapter focuses on the structural features of RNA and the attempts made to target RNA previously. The second chapter focuses on the development of a small molecule library enriched with substructures derived from RNA binding ligands. For this study a fragment-based approach (fragment based approach is detailed in chapter 2) is used in order to accommodate the conformational flexibility of RNA. The library molecules are used for screening against suitable RNA targets using NMR. We identified at least 5 ligands out of which 2 are novel ligands binding to the ribosomal 16s rRNA.

The third chapter is focused on the role of small molecules in inducing conformational changes in an RNA genetic regulatory element called the S-Adenosyl methionine (SAM) SAM-I riboswitch. The mechanistic features of the SAM-I riboswitch to understand the basis for specificity and discrimination and its gene regulation mechanism are reported. To address the conformational dynamics Bacillus subtilis and Thermoanearobacter tencongensis SAM-I riboswitches in response to SAM binding several conformer mimics are designed, synthesized and characterized using NMR, equilibrium dialysis, and inline probing. The study shows that apart from the conserved residues of the binding pocket, residues downstream
of the binding pocket are involved in detecting SAM and assist the binding of SAM to the riboswitch with weak affinity.

Our data highlights the capacity of a so-called antiterminator helix from the expression platform to assist the formation of a partial P1 helix of the aptamer domain. A stable P1 is involved in recognition and tight binding of SAM. Our in vitro experiments suggest that the riboswitch could switch from an unbound conformation to tightly SAM bound structure through weakly binding intermediate structures in the presence of the small molecule SAM. The future directions are included in the fourth chapter along with the conclusions.
CHAPTER 1: INTRODUCTION

1.1. RNA as a Potential Small Molecule Drug Target

Until recent times, RNA was mostly considered as a passive messenger in cellular protein synthesis. New RNA structures with hitherto unknown functions are projecting RNA as an active element with significant roles in enzyme-like catalysis [1], gene regulation [2] transcriptional control, and translational control [3-5]. Recently non-coding RNAs (ncRNA) in prokaryotes and eukaryotes have been shown to be involved in gene regulatory function. In eukaryotes, evidence is established for roles of ncRNA in epigenetic memory, physiology and development and also in chromatin architecture [6]. These recent discoveries conspicuously indicate RNA as a potential drug target for various diseases. In some pathogenic viruses, RNA is the genetic material and plays a crucial role in replication. Various RNA structures within the human immunodeficiency virus (HIV-I) genome interact with host proteins [7]. Such RNA structures also prove RNA as a potential drug target. It could be more effective to attack such viruses at RNA level and protein level since the virus utilizes host protein machinery to synthesize its proteins (strategies involving combination therapies).

Proteins and DNA are extensively explored drug targets. Out of the numerous thousands of cellular proteins, nearly 1620 have been identified to be involved in genetic diseases and out of these, only about 207 proteins have been targeted within humans [8]. It is also shown by other analysis that only about 15% of the total proteome is accessible for drug targeting [9]. Success with protein targeting has been largely with specific protein classes such as kinases, neuraminidases, GPCRs (G protein-coupled receptor), nuclear receptors, cytochrome proteins and ligand gated ion channels [8]. The common protein targets used
for drug discovery are enzymes with compact active sites. One of the difficult challenges in targeting some large proteins is to deal with a large interfacial area involved in the interaction (particularly in protein-protein, protein-DNA interactions) required for specific recognition. The interfacial area is recognizable through surface polarity and distribution of charge [10]. Principles for drug discovery targeting DNA have been established based on the understanding of the interactions with natural ligands [11, 12]. Such ligands have less druggable properties because of issues with solubility and nuclear membrane permeability and toxicity [13]. However, the drug-DNA interactions are less in number as compared to the Drug-protein interactions but have more successful stories in drug discovery. Since RNA is similar structurally to DNA and functionally (folding) to protein, the aforementioned considerations while targeting RNA should be considered.

In most cases, the function of the RNA is dependent upon binding of either a small molecule or a protein or a small RNA to the functional RNA. Such binding events are potential points of intervention for drug targeting. Though RNA is much simpler than proteins in basic constitution, its structural properties allow it to create pockets for ligand binding. Small molecules can counteract the long range tertiary interactions constituted by hydrogen bonds and Van der Waals and weak electrostatic forces between helices in the secondary structure of the RNA[14]. Other structural features of RNA like non Watson-Crick base pairing, coaxial stacking and, bulges etc., are significant in binding to small molecules [15, 16]. However, RNA based small molecule drug targeting did not reach to the point of having established principles and methodology as compared to proteins and DNA based drug discovery. The folding of RNA and small molecule interactions have only recently been the subject of focused study.
1.2. RNA Structural Features and Their Contribution towards Small Molecule Binding

1.2.1. RNA Structural Hierarchy

Commonly, four repeating nucleotide residues comprise the basic sequence of RNA unlike proteins, which are comprised of 20-plus amino acids. Albeit like proteins, RNA has structural hierarchy with primary, secondary and tertiary structures. The primary structure is comprised of the nucleotide sequence. Helices, bulges and internal loops formed by unpaired nucleotide residues, hairpin loops, junctions and branch points form the secondary structure of RNA. The tertiary structure of RNA is formed by weak interactions such as Van der Waals interactions, weak hydrogen bonds, and electrostatic interactions between residues of different secondary structure elements. Distant tertiary contacts are mediated through metal ions, anions, or small molecules [17-19]. The amount of energy required to stabilize the tertiary structure is less than that required for secondary structure stabilization contributed from Watson-Crick base pairing [20]. Table 1.1 summarizes the energies required for the stabilization of RNA secondary structures.

1.2.2. RNA Secondary Structure in Detail

The dominant representatives of the RNA secondary structures are helices. The Watson-Crick base pairs are the dominant base pairs present in the RNA double helix structures. The double helix adopted by RNA is usually an A-form helix with 11 base pairs per turn. The A-form helix is shorter and wider than the B-form helix, which predominates in DNA. The major groove is deep and narrow and the minor groove is shallow to allow the close packing of the double helical and single stranded RNA for function [22-24] (Figure 1.1 a).
Table 1.1: The thermodynamics of the RNA secondary structures. “n” is the number of bases [20, 21].

<table>
<thead>
<tr>
<th>Secondary structure element</th>
<th>$\Delta G^0$ (kcal mol$^{-1}$)</th>
<th>$\Delta H^0$ (kcal mol$^{-1}$)</th>
<th>$\Delta S^0$ (cal mol$^{-1}$)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hairpin closure</td>
<td>+4.5 to +5.9</td>
<td>0</td>
<td>-14.5 to -19.0</td>
</tr>
<tr>
<td>Internal loop closure</td>
<td>+4.1 to +6.3</td>
<td>0</td>
<td>-13.2 to -20.3</td>
</tr>
<tr>
<td>Bulge loop closure</td>
<td>+3.9 to +5.5</td>
<td>0</td>
<td>-12.6 to -17.7</td>
</tr>
<tr>
<td>Watson-Crick base-pair</td>
<td>-0.9 to -3.4</td>
<td>-5.7 to -14.2</td>
<td>-15.5 to -34.9</td>
</tr>
<tr>
<td></td>
<td>Average=-19</td>
<td>Average=-9.6</td>
<td>Average=-24.9</td>
</tr>
</tbody>
</table>

The sugar ring in nucleic acids is inherently nonplanar. This nonplanarity is termed puckering. The configuration of the non-planar sugar ring in the nucleotides is determined by the sugar-pucker phase angle and the sugar puckering amplitude. The sugar ring can be specified by the five-endocyclic torsion angles (Figure 1.1c). The sugar-pucker phase angle and amplitude are defined in terms of the five torsion angles. The most common configurations of the sugar pucker are C2’endo and C3’ endo (Figure 1.1b). The C2’ and C3’ are carbon positions in the sugar ring. The endo or exo describes the orientation of the carbon towards the plane (facing the 5’OH of the sugar) or outside the plane respectively. The 2’ OH of the ribose sugar in RNA contributes to the C3’ endo conformation, inhibiting the formation of C2’ endo conformation in-turn resulting in the A-form geometry [22]. The deep and narrow major groove does not allow the formation of binding pockets in classic A-form helices. The major groove is information rich because the functional groups of the nucleotide bases are projected into the major groove, resulting in a varied electrostatic landscape [23]. The major groove widens upon binding by protein, peptide, or small molecules like deoxystreptamine [15, 25, 26]. This major
groove widening is favored near non-paired bulges and hairpins [22]. For these reasons, interacting proteins or peptides can achieve specific recognition by contacting the major groove near non-paired regions. Figure 1.2 demonstrates the secondary structure schematics of some of these unpaired secondary structure motifs a) Duplex, b) Bulge loops that cannot base pair on one side of the structure, c) Interior loops that cannot form base pairs on both the sides of the RNA, d) hairpin loop e) mismatch f) symmetric internal loop g) asymmetric internal loop h) Junctions in which two or more helices integrate to form a single structure.

Figure 1.1: A form and B form helices (left), sugar pucker configurations (right) in nucleic acids [27]. Sugar pucker (bottom). (Figure Adapted from [[27-29]. The cartoons for the A-form (nucleotide database ID: 2KUW) and B-form (nucleotide database ID: BDL018) helix are generated using RCSB Jmol viewer and the biological assembly generated using PISA software). A cartoon representation of a RNA double helix showing the minor groove with the sugar moieties and the nitrogen bases facing the major groove (figure generated using mc sym pipeline, pymol).
1.2.3. RNA Secondary Structure Is a Rational Target for Small Molecules

As mentioned earlier, the RNA major groove has varying electronegative potential. This arises because of the phosphate groups of the nucleotide residues but also from oxygen and nitrogen moieties on the bases.

![Figure 1.2: Schematics of the secondary structure elements that are frequently found in RNA. A) duplex b) bulge c) single stranded RNA between two hairpins d) hairpin e) mismatch f) symmetric internal loop g) asymmetric internal loop h)Three stem i) four stem (The three and the four stem show the junction) j) Single nucleotide bulge.](image)

The functional groups of the nucleo-bases face the major groove resulting in the projection of the 2’OH groups of the residues into the minor groove. Various RNA structures like the ribozymes [30, 31], HIV-1 RRE [32], and A-site rRNA bind to ligands near bulge regions or stem loops. Some small molecules are involved in minor groove interactions with specific 2’OH hydrogen bonding [33].

The energy required for the formation of the tertiary structure in RNA is lower than that required for the formation of the secondary structure. The latter is strongly stabilized by base pairing and by stacking interactions between each base pair. Hence perturbation of the tertiary structure to prevent the interactions between independent secondary structures is the key to manipulating RNA structure and function with small molecule
ligands. The riboswitches (detailed in later sections) present a new class of novel RNA targets in which small molecule ligand binding perturbs the RNA secondary structure.

### 1.3. Molecular RNA Targets

Several molecular RNA targets with crucial biological functions have been identified and shown to bind to various small molecule ligands to date. Some of the targets include bacterial ribosomal RNA, bacterial T-box RNA, bacterial tRNA, Hepatitis C Virus Internal ribosome entry site (IRES) RNA, Human immunodeficiency viral (HIV) Trans activating response element (TAR) RNA, Rev Response element (RRE) RNA, mRNA structures that include riboswitches, and 3’ untranslated regions [34-37]. In all these examples, the efficacy of the small molecule depends on its ability to prevent the RNA-protein, RNA-RNA, RNA-DNA interactions. The intervening small molecule binding may block the biological function of the RNA.

### 1.4. Small Molecule RNA Interactions: Insights from Aminoglycosides

Amongst the various small molecule species the Aminoglycosides represent the species of small molecules that have been studied in complex with RNA the most extensively. The following sections will briefly summarize lessons learnt from the RNA-Aminoglycoside interactions.

#### 1.4.1. Amine groups of the Aminoglycosides Are Responsible for the Electrostatic Interactions between Small Molecules and RNA

The amine groups of the Aminoglycosides recognize the phosphate backbone of the RNA [31, 32]. The binding affinity of the small molecule is also dependent on the number of
amines present on the small molecule. The electronegative charge spectrum created by the phosphate backbone, and oxygen and nitrogen moieties of the bases of the RNA in the major groove creates a charge contour for small molecule recognition and binding similar to that of protein surface charge spectrum. Altering the pH in vitro altered the binding affinity of certain Aminoglycosides [38, 39]. By introducing more electro positively charged groups (Guanidinium groups) in the place of amine groups in the Aminoglycosides, increased affinity is observed to RNA [40-42]. The significance of the electrostatic interactions are also demonstrated using varying salt concentrations in binding assays [41]. By varying the NaCl concentration, the binding affinity could be increased or decreased. Further thermodynamic studies of aminoglycoside binding with ribosomal RNA in humans and E.coli show that more than half of the binding energy is constituted from electrostatic interactions [40, 43]. The electrostatic interactions between RNA and small molecule contribute to high affinity binding. However one disadvantage with electrostatic interaction is that a number of non-specific interactions with many RNA targets leading to promiscuous RNA binding are observed with many Aminoglycosides. However the specificity of the Aminoglycosides could be increased by increasing the salt concentrations in the binding assay to eliminate weak electrostatic interactions [44].

1.4.2. Nonionic Interactions

Though the Aminoglycosides are positively charged and the RNA is negatively charged, various nonionic interactions such as stacking and packing interactions stabilize the RNA-Aminoglycoside complex. Several Aminoglycosides make nonionic interactions that bury
the non-polar moieties of the small molecule into the hydrophobic groove of the RNA. It is observed from 16s rRNA A-site construct, that ring I of the Gentamycin is involved in stacking interaction with the G1491 and A1492 residues of the RNA [45].

1.4.2. Pseudo Base Pairs Facilitate Small Molecule Recognition

RNA is highly structured as mentioned in the previous sections. It consists of A-form double helices with intervening bulges, and loops resulting from the non-canonical/non Watson-Crick base pairs. A close look at the aminoglycoside bound-RNA complexes, show that the rings of the Aminoglycosides stabilize such unpaired/non-canonical residues. The stabilizing interactions could be the mode of recognition in the binding sites of the RNA. This recognition mode is well studied through NMR and Circular Dichroism for an A-site RNA construct (a specific construct designed and synthesized for in vitro studies) with the Deoxystreptamine Aminoglycosides such as Apramycin [47]. Westhof and colleagues have shown high-resolution structures for the A-site RNA in complex with the Aminoglycosides Gentamycin and Paromomycin [46, 48]. They have done extensive work on the role of the base pairs and pseudobasepairs in ligand recognition.

1.4.4. Water Mediated Contacts

The 2’OH group of the ribose sugar in RNA is a strong hydrogen bond donor. It contributes to the excessive hydration of RNA in the minor groove by forming a highly ordered water network [49]. The excess water molecules in various ligand interacting regions of the RNA mediate ligand binding.
Figure 1.3: A cartoon representation of the Gentamicin C1A and the A-site RNA fragment. Orientation of Gentamicin C1A (ball and stick representation) molecule inside the A site RNA duplex (pink cartoon). Ring numbers of Gentamycin C1a are designated I–III. (Figure using adapted from [46] and generated with crystal structure of NDB ID: DR0015 with RCSB Jmol viewer).

The water molecules rearrange themselves and optimize the hydrogen bond distances between the small molecules and the RNA. The Aminoglycosides are well explored in the context of water mediated hydrogen bond formation in the 16s A-Site ribosomal RNA [40, 44]. Such studies establish that the water molecules also have a role in compensating the energy penalty of dehydration within the RNA binding pocket upon tight binding of the ligand [50].
1.4.3. Conformational Flexibility of RNA Accounts for Small Molecule Binding and Attaining Shape Complementarity

A small subset of ligand bound RNA structures show that RNA allows the recognition of the small molecule ligand and by conformational adaptation, the RNA attains high binding affinity with the ligand. The conformational flexibility of RNA is a bane for RNA based drug discovery, as it is problematic for structure based drug design [51]. From the lessons learned from the hammerhead ribozyme-aminoglycoside binding and A-Site-aminoglycoside binding, it is hypothesized that the RNA conformationally adapts to a structure that attains high affinity with the ligand [52-54].

1.5. Riboswitches Are Classic RNA Systems to Understand Diverse Small Molecule-RNA Interactions

From the above literature review focused on identifying small molecule ligands binding to RNA, the two reviews by Aboul-Ela [55] and Hergenrother [56] that comprehensively summarize the various small molecules developed to target RNA, it is evident that most of the RNA binders (or at least the most frequently studied) are Aminoglycosides, Chloramphenicols, Tetracyclines, and DNA binding intercalators. As discussed above, Aminoglycosides and their derivatives are nonspecific as observed by their ability to bind to many RNA targets. Even derivatization of the Aminoglycosides has not resulted in achieving high specificity and affinity in most cases [57, 58]. Other major drawbacks in using Aminoglycosides as drug molecules are their cytotoxicity at high dosages (for attaining high affinity with the target RNA) and problems with drug delivery.

The recent discovery of regulatory mRNA elements in the non-coding regions called the ‘Riboswitches’ has opened venues for understanding RNA-small molecule interactions
and RNA based drug discovery. New riboswitches have been discovered continuously during the last decade. Each of these riboswitches recognize their cognate ligand with high specificity and bind with high affinity [59]. The ligands that interact with the riboswitches are structurally diverse. The mechanisms by which the riboswitches interact with their ligands vary from one riboswitch to another [59, 60].

Riboswitches are functionally significant for their gene regulatory mechanisms. They are gene regulatory elements that function at the translational or transcriptional levels and that work with minimal gene regulatory machinery[61]. The riboswitches operate within two domains. The aptamer domain binds to the ligand and the expression platform undergoes structural changes upon ligand binding to modulate gene expression[61, 62]. High sequence specificity is found in the ligand-binding domain in classes of riboswitches that bind to the same ligand. For example there is sequence conservation between the SAM-I,II,IV riboswitch classes. All the three classes bind to the small molecule ligand S-Adenosyl Methionine (SAM). The non-conserved expression platform signifies that though some riboswitches recognize similar ligands, they can operate by different mechanisms. The sequence conservation in the ligand-binding domain is shown to be responsible for the specific recognition of the cognate ligands. An in-depth understanding of small molecule interactions with riboswitches, increases the scope for developing more novel and structurally diverse RNA targeting drugs[63]. The structural and mechanical studies that decipher the reasons for the riboswitch specificity to its cognate ligand and the ability of the small molecule to alter the riboswitch structure upon binding will contribute significantly towards RNA based drug development. The following few sections
will describe aspects of the riboswitches including structures of ligand binding domains and their gene expression controlling mechanisms.

1.5.1. Riboswitches in Gene Expression Regulation

The process of gene expression includes the transcription of DNA into messenger RNA (mRNA). The information on the mRNA is then translated into protein. Various gene expression control mechanisms have been discovered at transcriptional and translational levels. These mechanisms involve DNA-protein, DNA-DNA, DNA-RNA, RNA-Protein interactions. The role of RNA in gene regulation as attenuators in E.coli trp operon located near the 5’ untranslated region is shown previously [64]. The non-coding RNAs (ncRNA) operate by either interacting with coding mRNA through complementarity or by interacting with gene regulatory proteins [65, 66]. The riboswitch ability to modulate gene expression gained attention since the last decade since Henkin and coworkers identified gene regulatory elements in the 5’ untranslated regions of gram-positive bacteria. These regulatory mRNA elements called the riboswitches operate at transcriptional and translational levels by interacting with small molecules without accessory proteins or DNA [61, 67]. The riboswitches have high selectivity for their cognate ligands [68, 69]. They undergo structural and functional switching upon ligand binding.

In transcriptional riboswitches the expression platform undergoes structural rearrangement upon ligand binding. The conformational changes lead to the stabilization or destabilization of a transcription terminator hairpin downstream of the ligand binding site [70]. In the presence of the ligand the terminator helix is stabilized. This
configuration is referred to as the ‘OFF’ state. In the ligand free state, the RNA folds to form an alternative helix called the antiterminator (AT) that prevents the formation of the intrinsic terminator. The RNA with AT is referred to as the ‘On’ state. Some riboswitches work by positive feedback regulation. The binding of the ligand in these riboswitches will stabilize the On state. The interplay of these mutually exclusive terminator and antiterminator helices in the presence or absence of the ligand results in gene expression control.

During translation of the mRNA, the Shine-Dalgarno (SD) sequence on the mRNA is necessary for translation initiation by ribosomal recruitment onto the mRNA. In translational riboswitches, the ligand binding to the aptamer leads to sequestration of the SD sequence via the formation of anti-SD helix (ASD). In the “On” state, the formation of the ASD is prevented by an alternative helix anti-anti-SD (AASD) [71, 72].

A third category of riboswitch mechanism is through splicing. Ligands binding to the ribozyme based riboswitches undergo structural changes to stabilize a ribozyme in the catalytic state. The ribozyme cleaves the mRNA preventing further transcription/translation [73, 74]. Schematic depiction of the transcription, translation and ribozyme based riboswitch mechanisms are shown in Figure 1.4.

1.5.2. Classification and Distribution of Riboswitches

The Riboswitches respond to a variety of biological and physical stimuli that range from temperature to metabolite and metal ion concentrations in the cellular environment. Most effector molecules are metabolites such as amino acids, nucleic acids, vitamins, enzyme cofactors and metal ions. The metabolite effectors of riboswitches in most cases
are the end products of cellular metabolic pathways. Hence the riboswitches are involved in gene regulation based on a feedback mechanism that can result in down-regulation or up-regulation of genes involved in the metabolite synthesis. Sequence conservation is limited to the aptamer domain in riboswitches belonging to the same class. However, more than sequence conservation, the secondary structural elements are conserved in the aptamer domains responsible for ligand recognition [75]. Riboswitches are classified as families and classes. The classification is based mainly on the secondary structural elements and the effector ligands or the signal to which the riboswitch responds.

For example, the SAM riboswitch family (All the riboswitches that detect SAM fall into this family) has 5 classes (SAM riboswitches with conserved secondary structural elements fall into a designated class). To date, 24 classes of the riboswitches are identified with autonomous gene control mechanisms in various forms of life [76].

The riboswitches based on mechanism of gene regulation and location of the expression platform and aptamer are categorized into five categories. The five categories include antisense regulation, transcription regulation, translation regulation, dual transcription and translation regulation, and direct translation regulation [59].

In the case of SAM riboswitches, there are five classes. The secondary structural elements required for SAM recognition are conserved more than the sequence of the aptamer domains in each class. Less sequence but more secondary structure conservation is seen in the expression platforms because the riboswitches that bind to the same metabolite (SAM riboswitches) have different modes of gene regulation at transcriptional or translational levels. Such diverse mode of operation of the riboswitches in response to the
same metabolite indicates shared origin but independent evolution of individual riboswitch classes [77].

**Figure 1.4:** Schematic representation of riboswitch mechanisms. The red circle indicates the metabolite molecule. RBS is the ribosome-binding site. The single arrow indicates the cleavage point in ribozyme control.

Several methods have been implemented in identifying the prevalence of riboswitches in various organisms. The search for known riboswitches and new riboswitches in various organisms is complicated when using simple genomic database searches because of the sequence variability in the expression platform and aptamers as mentioned earlier.
Identification of riboswitches in the 5’ untranslated region of mRNA has been implemented using comparative genomics, specific sequence elements, consensus secondary structure elements and covariance models [59, 78]. Most riboswitches discovered to date are found in bacteria. Some riboswitches are found in all three forms of life [79, 80]. The riboswitch studies from metagenomics indicate that the THI box, B12, and glycine riboswitches are the most abundant riboswitches followed by SAM and FMN riboswitches while glmS, lysine, and purine riboswitches representing the least abundant. Very few riboswitches have so far been identified in the eukaryotes.

1.5.3. Non Cognate Riboswitch Small Molecule Ligands for Antibacterial Drugs

As mentioned in the earlier sections, nearly 24 classes of riboswitch are discovered to date and most of them are found in bacteria. Riboswitches are novel chemical scaffolds for antibacterial drug targets with their distribution in such vast a spectrum of microorganisms and their involvement in potential gene regulation of essential metabolites. The Figure 1.5 [63] summarizes the gene control by riboswitches in various pathogenic bacteria.

The presence of different riboswitch classes in such a variety of bacteria opens venues for developing novel RNA binding antibacterial drugs. The antibacterial activity by targeting riboswitches can be achieved largely by 1) Developing ligands that compete with the cognate ligand and prevent the binding of the latter to fix the riboswitch either in the ‘On’
that may be riboswitches have been identified additional classes of conserved RNA motifs recognize SAM. A star denotes a riboswitch functional group. Three different structural they bind. The expected protonation states at by that riboswitch is predicted to be essential indicate that at least one of the genes regulated in parentheses by the total number of genes of representatives of each riboswitch class pathogens that carry riboswitches. The number (second subdomain that coordinates two metal receptor interfaces and show equally binding riboswitch from
forms a three-dimensional (3D) 22. Members of other riboswitch classes form similarly intricate ligand-
), right panel) The distribution of known riboswitch
indicates the number of genes regulated, red color signifies that the genes regulated by those riboswitches are responsible for virulence of the bacteria (Adapted from [63]).

Figure 3b: The distribution of riboswitches in pathogenic bacterial species. The numbers indicate the representatives of each class, the number in parenthesis indicates the number of genes regulated, red color signifies that the genes regulated by those riboswitches are responsible for virulence of the bacteria (Adapted from [63]).

or ‘Off’ configuration (based on the mechanism of the riboswitch), 2) by developing ligands that bind sites other than the aptamer site and preventing the cognate ligand binding in a non competitive manner or sequestering the elements in the expression platform required for the formation of decision making helices (AT,AAT,SD, ASD, AASD etc).

From Figure 1.5 it is evident that the Thiamine pyrophosphate (TPP) riboswitch, Flavin mononucleotide riboswitch (FMN), Purine riboswitch, Adocobalamine (AdoCbl) riboswitch and the SAM-I riboswitch are prevalent in most pathogenic bacteria. It is also noteworthy that the some of the genes that are regulated by these riboswitches in these bacteria are responsible for the virulence of these bacteria. Several methods that have
been adapted from general drug discovery have been used to identify novel ligands that bind to these riboswitches.

1.5.3.1. FMN Riboswitch Ligands

The FMN riboswitch is found in the 5'UTRs of gram-positive bacteria. It regulates the expression of genes involved in riboflavin metabolism. FMN is a phosphorylated derivative of riboflavin. Inline probing studies have revealed that the FMN riboswitch, also called the RFN element, controls gene expression by transcription/translation termination [81, 82].

A naturally occurring pigment, roseoflavin binds to the FMN riboswitch. [81]. Inline probing analysis (Method described in chapter 3) of the ribD FMN riboswitch indicates that roseoflavin binds with an “apparent K_d” of ~100nM as compared to its cognate ligand FMN of 3nM and riboflavin of 5nM. (The affinity is mentioned as apparent K_d in some references because the measured value is not the actual K_d, it is the K_d observed using a particular method, in this case inline probing. Different K_d values are reported for the same SAM riboswitch construct using different techniques (table 3.3 in chapter 3). However some authors report apparent K_d as K_d). The phosphate groups of the FMN render tighter binding with the aptamer. The ribityl groups of the riboflavin and roseoflavin bind differently than the FMN ribityl group. Mutational analysis and structural probing data suggest that the aptamer adapts to the ribityl groups of the riboflavin and roseoflavin during binding. Such binding with riboflavin and roseoflavin is observed in FMN riboswitches of more than one bacterial species [83]. The role of the FMN riboswitch
is in agreement with the antibacterial activity of roseoflavin from the experimental data mentioned above.

1.5.3.2. Lysine Analogs Targeting Lysine Riboswitch

The lysine riboswitch is found in many gram positive and gram-negative bacteria including pathogenic *E.coli* and *B.subtilis*. The riboswitch controls the expression of the *lysC* gene that codes for aspartasekinase II enzyme responsible for lysine synthesis [84]. It also controls the expression of other genes such as *yvsH* involved in the transport of L-lysine. The lysine riboswitch aptamer senses the L-lysine and modulates gene expression by transcription termination. The lysine riboswitch secondary structure elements consists of the helices P1-P6, junction regions and loop E involved in interhelical interactions and Kink-turn (Loop E is an internal symmetric loop with non-canonical base pairs, kink turn is a sharp bend or kink between two double-stranded elements and it bends toward the minor/shallow groove [85-87]) formation [88]. The secondary structure undergoes reorganization upon ligand binding [89, 90]. The ligand-binding site is within the P1-P5 helices. The P1, P2 and P4, P5 helices are stacked with the inter-helical tertiary interactions between loop regions mediated by a kink-turn motif in the P2 helix [91]. The crystal structure of the lysine riboswitch aptamer in complex with lysine revealed the structural elements required for the lysine recognition. Charges associated with the two amino groups and a carboxyl group at the two ends of Lysine are recognized by the aptamer. The high structural organization of the binding pocket for lysine in the aptamer domain ensures high discrimination of lysine against its metabolite analogs [91].
The Figure 1.13 illustrates the binding pocket interactions of the lysine riboswitch with lysine from the crystal structure.

Figure 1.6: Lysine analogs that were designed to target the lysine riboswitch (The shaded regions are the modified functional groups) (Figure adapted from [88, 91]).

In spite of the lysine riboswitch's high selectivity for lysine, several antibacterial lysine analogs such as L-4-Oxalysine, L-Homoarginine, N-6-Acetyl-L-lysine, N2-Methyl-L-lysine, N6-Trimethyl-L-lysine, 3-Amino-L-tyrosine are shown to bind to the lysine riboswitch. These analogs not only bind to the riboswitch but also decrease the growth of the bacteria invivo [88]. Further examination of the binding pocket revealed that a cavity in the binding pocket allows the binding of the lysine analogs. Of all the analogs, the S-(2-aminoethyl)- L-cysteine (AEC) binds with 30 fold less affinity as compared to L-Lysine. The AEC modulates the structure of the riboswitch at high concentration. It is also shown that the antibacterial resistance against these analogs is obtained by altering structural elements involved in aptamer tertiary interactions[92]. In a research conducted by
Breaker et al, they observed that the growth of *B.anthracis* and *B.subtilis* in lysine rich media, in the presence of the above ligands, is not inhibited [88]. This indicates that some ion transporters assist the bacteria in using the lysine from the media [93]. In such a case, though the genes responsible for lysine biosynthesis are repressed, through targeting the lysine riboswitch, potential bacterial growth inhibition cannot be attained if the bacteria can depend on host lysine alternatively.

1.5.3.3. Small Molecules That Target TPP Riboswitches

The Thiamine pyrophosphate (TPP) class of riboswitches are involved in the regulation of genes involved in thiamine metabolism and transport. Thiamine is a crucial cofactor of enzymes involved in catalysis of glycolytic, citric acid cycle, and pentose pathway reactions. Intracellular thiamine is phosphorylated to TPP. TPP participates in the aforementioned reactions to cleave carbon-carbon bonds adjacent to carbonyl groups. The TPP riboswitch is found in the 5'UTRs in bacteria, plants and fungi. It binds to the TPP specifically and alters the gene expression of the downstream genes *thiM* and *thiC* involved in Thiamine metabolism [94]. The modes of operation for TPP riboswitches vary in different organisms. In certain bacteria, it operates at the translational level by sequestering the SD sequence. In fungi, a ribozyme mediated gene control by the TPP riboswitch that controls the mRNA splicing is seen [74]. Inline probing studies with the *thiC* and *thiM* TPP riboswitches have indicated that the phosphate groups and the amino group make specific contacts with the TPP riboswitch aptamer (Figure 1.15).

Equilibrium dialysis studies and inline probing studies with the TPP riboswitch show that various analogs of TPP such as Benfotiamine, Amprolium, Oxythiamine and Thiamine
disulfide make weak contacts with the \textit{thiM} and \textit{thiC} riboswitch. However these analogs though similar to TPP, fail to modulate the overall riboswitch structure upon binding\cite{94}. In the same research, the authors also observed that the TPP derivatives, Thiamine phosphate and Thiamine bind to the \textit{thiC} and \textit{thiM} riboswitch and modulate the riboswitch structure as observed by the inline probing studies. In several instances, the isosteric Thiamine analog Pyrithiamine (PT) is shown to inhibit the growth of many bacteria and fungi and in eukaryotes, PT is phosphorylated to PTPP like Thiamine and PTPP inhibits the role of TPP in enzyme catalysis\cite{95,96}. In other microbial studies, it is seen that the expression of genes involved in the thiamine metabolism like the \textit{thiA}, \textit{tenA} are altered when certain fungi and bacteria are grown in media containing PT. The binding of PT, PTPP and thiamine to the TPP riboswitches was studied recently to probe the involvement of the TPP riboswitches in the antibacterial and antifungal activity as mentioned above\cite{97}.

The Inline probing analysis with the \textit{tenA} bacterial TPP riboswitch shows that the PT and Thiamine and PTPP significantly bind to the riboswitch with an affinity of \(~6\mu M, 50\mu M, 160nM\) respectively as compared to 50nM with TPP. These compounds are shown to have antibacterial activity and PT resistant bacteria are shown to undergo mutations in the TPP riboswitch aptamer\cite{97}. The \textit{thiA} TPP riboswitch in the \textit{Aspergillus oryzae} fungi also is shown to bind to the PTPP with an affinity of 56nM as compared to 50nM with TPP. It induces structural changes to alter gene expression control in fungi \cite{97}. 


1.5.3.4. Purine Analogs That Bind the Guanine Responsive Riboswitch

The Guanine responsive riboswitch controls the expression of genes involved in purine metabolism and transport. The secondary structure of the aptamer domain of the Guanine riboswitch consists of helices P1,P2,P3 and loops L2 and L3. The tertiary interactions between the helices are mediated through a pseudoknot motif. The L2 and L3 loops are pre-structured and assist the aptamer in ligand recognition. Upon ligand recognition, the three helices bind tightly by completely enveloping the ligand [98, 99]. A downstream terminator helix is formed upon ligand binding. In the absence of the ligand an AT helix is formed and prevents formation of the terminator. Sequence comparison of four common Guanine riboswitches suggested that common Guanine analogs could be used to target all the Guanine riboswitches [100]. The crystal structure of the Guanine bound Guanine riboswitch shows that a Watson-Crick base pair in the aptamer is crucial for Guanine binding and also the examination of two regions in the aptamer revealed that additional groups could be added in the C2 and the C6 region of the aptamer without losing binding affinity. Sixteen purine analogues have been designed rationally to bind to the Guanine riboswitch. Inline probing analysis, mutational analysis and invivo studies indicate that the G7, G6 and G15 compounds bind with significant affinity to two different bacterial guanine riboswitches. The G7 compound shows antibacterial activity [100, 101]. Four novel purine riboswitch ligands are obtained by Brenk and group using RNA-ligand docking [101].
Figure 1.7: The ligands designed for targeting the Guanine riboswitch with apparent $K_D$ values with the guanine riboswitch. (Figure adapted from [102]).

1.5.3.5. T-Box RNA and Its Ligands

The T Box RNA is one such regulatory RNA in the gram-positive bacteria. The T Box RNA is found in the 5’ untranslated region of the mRNA in *B.Subtilis tyrS* gene, encoding tyrosyl-tRNA synthetase. It is involved in regulating the amino acyl tRNA synthetases [103]. The regulatory effect is a result of the conformational alteration of the mRNA to form either a terminator or an antiterminator helix in the presence or absence of charged tRNA respectively. The anticodon loop of the tRNA interacts with the 5’ region stem loop of the T Box RNA. The 5’ portion of the T Box detects the presence of the amino acid in the charged tRNA. In the absence of the charged tRNA, the 3’ end of the uncharged tRNA interacts with the 3’ region of the T Box to form a stable antiterminator that allows the synthesis of proteins necessary for charging of the tRNAs. The bulge of the antiterminator

<table>
<thead>
<tr>
<th>R</th>
<th>$K_D$ (nM)</th>
<th>R</th>
<th>$K_D$ (nM)</th>
</tr>
</thead>
<tbody>
<tr>
<td>(G1) $\text{CF}_3$</td>
<td>2000</td>
<td>(G4) $\text{NCH}_3$</td>
<td>0.5</td>
</tr>
<tr>
<td>(G2) $\text{NCH}_3$</td>
<td>200</td>
<td>(G5) $\text{NCH}_3$</td>
<td>500</td>
</tr>
<tr>
<td>(G3) $\text{NH}_2$</td>
<td>100</td>
<td>(G6) $\text{NH}_2$</td>
<td>3300</td>
</tr>
<tr>
<td>R'</td>
<td>$K_D$ (nM)</td>
<td>R'</td>
<td>$K_D$ (nM)</td>
</tr>
<tr>
<td>(G7) $\text{OH}$</td>
<td>20</td>
<td>(G12) $\text{H}_2\text{NCH}_3$</td>
<td>80</td>
</tr>
<tr>
<td>(G8) $\text{NCH}_3$</td>
<td>20</td>
<td>(G13) $\text{NCH}_3$</td>
<td>200</td>
</tr>
<tr>
<td>(G9) $\text{NH}_2$</td>
<td>200</td>
<td>(G14) $\text{NH}_2$</td>
<td>200</td>
</tr>
<tr>
<td>(G10) $\text{NCH}_3$</td>
<td>500</td>
<td>(G15) $\text{NCH}_3$</td>
<td>80</td>
</tr>
<tr>
<td>(G11) $\text{NH}_3$</td>
<td>8</td>
<td>(G16) $\text{NH}_3$</td>
<td>200</td>
</tr>
</tbody>
</table>
is conformationally free as calculated from NMR studies[104]. This region is logically amenable for targeting with small molecules and the 3’ end recognizing bases of the T Box also indicate sites for small molecule binding. Neomycin, streptomycin, oxazolidinones are a few examples of T Box antiterminator binding molecules [105].

1.5.3.5 Ligands Binding to S-Adenosyl Methionine (SAM) Riboswitch

The X-ray crystal structures of the SAM riboswitches, in recent years revealed the binding pocket at very high resolution. Five classes of SAM-I riboswitches (SAM-I to SAM-V) are identified to date and each class has its own conserved mechanism of SAM recognition and binding. Previously the SAM analogs S-Adenosyl Homocysteine (SAH) and S-Adenosyl Cysteine (SAC) have been shown to bind to the SAM-I riboswitch by isothermal calorimetry and equilibrium dialysis [106, 107]. From these experiments, it is shown that the SAM analogs can only bind to the SAM-I riboswitch with weak affinity, but cannot replace the SAM even at very high concentrations. Very recently a set of SAM analogs were shown to bind the SAM-II riboswitch. The authors chose SAM-II riboswitch after comparing the SAM interactions with the SAM-I and SAM-II riboswitches in the binding pocket. The SAM is tightly packed in a U conformation in the SAM-I aptamer with almost all the moieties of SAM making contacts with the aptamer residues [108]. In the SAM-II riboswitch, SAM is in linear configuration rendering the amino terminal and the carboxy terminal of the SAM some degree of freedom in the major groove of the aptamer. The authors speculated that analogs of SAM modified in the amino and carboxy terminals can potentially bind to the SAM-II riboswitch without loosing any affinity. Five SAM analogs
(Figure 1.17) have been tested and interestingly, analog 5 binds to the SAM-II aptamer tighter than SAM as seen by fluorescence binding experiments [109].

![SAM and SAH structures](image)

<table>
<thead>
<tr>
<th>R¹</th>
<th>R²</th>
<th>Compound</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH₂</td>
<td>OH</td>
<td>1</td>
</tr>
<tr>
<td>O</td>
<td>NH₂</td>
<td>2</td>
</tr>
<tr>
<td>O</td>
<td>NH₂</td>
<td>3</td>
</tr>
<tr>
<td>H</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td></td>
<td>OH</td>
<td>5</td>
</tr>
</tbody>
</table>

**Figure 1.8: S-Adenosyl Methionine and its analogs (Figure adapted from [109]).**

It is discussed in the earlier sections on RNA based drug discovery that promiscuity of RNA binding small molecule ligands is a concern to attain specificity and affinity. The riboswitches are novel RNA small molecule targets, since they are highly selective to their cognate ligands. The riboswitches are proved to be good antibacterial RNA drug targets as mentioned in the above sections. The approach to target the riboswitches mainly involves the characterization of the specific elements required for the recognition of the cognate ligand by the riboswitch. An in-depth understanding of the mechanism of the ligand recognition by the riboswitches and the discrimination by the ligand between the two states of the riboswitch (ON and Off states) may facilitate the design of ligands targeting riboswitches.
Our goal is to understand the RNA-small molecule interactions and develop methods to identify novel small molecules directed to RNA binding for RNA based drug development. The specific aims of my research are 1) To develop a RNA directed small molecule library (chapter 2) from the lessons learnt from aminoglycoside-RNA interactions and from Fragment based drug development strategies (since the RNA conformational flexibility and plasticity pose issues against the use of Structure based drug development (SBDD)) and screen against RNA targets, and 2) To probe the mechanism of the SAM-I riboswitch (chapter 3) with the goal of understanding the structural features of the riboswitch that contribute to the ligand specificity and to discrimination by SAM between SAM-I riboswitch gene expression-controlling conformers.

1.6. Concepts of Prokaryotic Transcription

Our work with the SAM-I riboswitch mentioned in chapter three is related to prokaryotic mRNA transcription. A brief review of prokaryotic transcription is introduced in this section. The bacterium *E.coli* is the organism in which mRNA, RNA polymerase have been first discovered [110]. The ground work for delineating the principles of the RNA transcription in Eukaryotes and prokaryotes started with *E.coli* transcription.

![RNA polymerase schematic in the closed-promoter-complex.](image)
The prokaryotic RNA polymerase is a polypeptide with four subunits α, β, β', and σ. The subunit σ is involved in the recognition of the transcription initiation sites. The core of the RNA polymerase consists of the subunits α(2), β(1) and β'(1) subunits. They catalyze the polymerization of the NTPs to RNA based on the information on the DNA strand.

Sequence comparisons of the genes in E.coli led to the identification of the transcription initiation sites called the promoters [111] The RNA polymerase σ subunit recognizes the promoter on the DNA and initiates polymerization. The nucleotide where the transcription is initiated on the DNA, called the start site, is designated +1. The promoter elements include two conserved six nucleotide sequences upstream of the start site at positions -10 and -35 [112]. Footprinting and mutation analysis at these two positions indicate that the σ subunit of the polymerase specifically binds to these sequences during transcription initiation. Lack of the σ subunit in the RNA polymerase leads to nonspecific binding by the enzyme [113, 114].

The RNA polymerase when first bound to the DNA is called the closed-promoter-complex. In a subsequent step, the RNA polymerase unwinds the DNA to form the Open-Promoter-Complex. The open-promoter-complex consists of the DNA with 15 unwound bases and the RNA polymerase. The DNA strand on which the RNA polymerase synthesizes RNA is called the template strand (Figure 1.1.9). The σ subunit is released from the core of the RNA polymerase after the first 10 NTPs are incorporated into the nascent RNA. After the release of the σ subunit, the polymerase moves from the initiation site to “elongate” the RNA with the elongation complex comprised of the RNA polymerase and DNA unwound to 17 base pairs. The elongation of the RNA occurs by the movement of the RNA
Polymerase along the template DNA strand, adding appropriate NTPs to the growing RNA strand in 5’-3’ direction.

The RNA transcription is terminated when the RNA polymerase reaches the transcription termination signal on the DNA. There are two common modes of prokaryotic transcription termination 1) Rho-dependent (Rho proteins are found in lower eukaryotes with function other than transcription termination [115]) and 2) Rho independent. The Rho-independent transcription termination depends on the formation of an intrinsic terminator loop [116]. The termination signal, when transcribed leads to the formation of RNA segment that forms a GC rich hairpin structure (terminator helix) followed by a U-rich segment. The stable GC rich terminator destabilizes the DNA-DNA dependent RNA polymerase interaction. The mechanical stress due to the stable terminator hairpin breaks the U-A bonds and releases the RNA strand and the RNA polymerase from the elongation complex. A proposed model for the release of the RNA strand form the polymerase is that the polymerase translocates forward without RNA synthesis [117], however, Nudler et al, have shown that the hairpin formation is the starting or the nucleation step for termination [118]. The hairpin then invades the RNA polymerase, causing RNA:DNA hybrid melting, structural changes of the catalytic site, and DNA-clamp (DNA clamp is a protein fold that assists the DNA dependent polymerase in catalytic activity by preventing the enzyme dissociation from the template strand) followed by the release of the RNA strand, and release of RNA polymerase without translocation.
CHAPTER 2: DESIGN AND IMPLEMENTATION OF AN RIBONUCLEIC ACID (RNA) DIRECTED FRAGMENT LIBRARY

2.1. Introduction

Drug development strategies targeting proteins rely on Structure Based Drug Design (SBDD) and High throughput Screening (HTS) in which millions of compounds are screened against a target to identify relatively strong binding molecules (medium and low throughput screening is done with small number of ligands). The SBDD is a strategy that is based on the use of a three dimensional target structure obtained by X-ray crystallography or NMR for drug development. In SBDD, a stable target structure with a defined binding pocket can be used to design small molecules specific to the binding pocket. In the case of RNA targeting, the flexible nature of RNA poses a challenge to the SBDD.

In recent years, Fragment Based Drug design along with structural biology resulted in viable clinical drugs targeting proteins [119, 120]. In fragment based approach “leadlike” molecules are obtained instead of “drug-like” molecules. Lipinski’s “rule of 5” (not more than 5 hydrogen bond donors (OH and NH groups), no more than 10 hydrogen bond acceptors (notably N and O), a molecular weight under 500 and CLogP (solubility of a molecule in a polar and non polar solvent) under 5 [121]) is traditionally accepted by researches in drug discovery as a criterion for drug like molecules. The leadlike molecules are smaller with a molecular weight of less than 300, and they contain smaller polar and hydrophobic surface areas than the drug-like molecules. A significant difference between leadlike and drug like molecules is, the lead like molecules are defined by the minimum limit of polar groups (Polar functionalities are necessary for chemically growing) the drug
like molecules are defined by the maximum limit of the polar groups. The “leadlike” molecules are more amenable for “derivatization” to pick up additional interactions with the binding site and thus increase the affinity to the target and still be drug like after derivatization.

Though 3d X-ray and NMR structures of many RNA elements are available, specific binders other than promiscuous Aminoglycosides have not been readily obtained, perhaps because of the flexibility of RNA and its plasticity. Few studies discovered novel RNA binding ligands through generic screens [122, 123]. In this context, the fragment-based strategy may be more useful in targeting RNA over SBDD and HTS. We adapted the principles from Fragment Based Drug Design and used cheminformatics tools to develop an RNA directed small molecule library to enhance the probability of binding to RNA. We developed the RNA directed library by identifying 120 RNA binding ligands reported in literature. A total of 114 fragments similar to the substructures of the RNA binding ligands from similarity search are incorporated into the library. The substructures are with molecular weight less than 300, with less than three hydrogen bond donors and acceptors and ClogP value 3. The fragments are then screened using NMR spectroscopy against a 27 nucleotide ribosomal A-site RNA that is shown to bind to aminoglycoside antibiotics. We obtained a set of hits out of which at least two of them are not reported as ribosomal A-site RNA binders.
2.2. Methods and Materials

2.2.1 RNA-Binding Ligand Database

The RNA binding ligands are identified from literature and included into the RNA-binding ligand database on the following criteria. We chose ligands that interact with RNA under a K$_d$ value of 50μM. In the case where a series of similar compounds as in Figure 1.8, the compounds with highest affinity are chosen along with the compounds that have distinctive functional groups (if they are within the 50μM K$_d$ range.

2.2.2. Fragmenting the Ligands from the Ligand Database In silico

The RNA binding ligands are fragmented using the software “chemdraw” Insilco to obtain substructures of the ligands. The fragmenting is done considering the rule 2 and rule 3 proposed by Schuffenhauer et al. The rule two states that for a given chemical structure, “Do Not Remove Rings with more than or equal to 12 Atoms if There Are Still Smaller Rings To Remove” and rule 3 states that “Choose the Parent Scaffold Having the Smallest Number of Acyclic Linker Bonds” [125]. The linker regions in the ligand structures are chosen as the point of fragmentation. We tried to retain the RNA binding moieties of the ligands in the substructures hence the remaining rules mentioned in the above reference are not considered. The fragmentation is also done to obtain substructures that fall within the “rule of 3” with molecular weight less than 250. In some cases where the polycyclic rings (e.g., tetracycline) did not have fragmentation points, the polycyclic ring was broken into components containing two to three rings each, but the original compound was also included as a fragment. Very large macrocycles, such as those occurring in some macrolides, were retained intact for the clustering analysis (section 2.2.3).
2.2.3. Clustering of the Fragment Database

The fragments obtained from the fragmenting of the RNA binding ligands are clustered into descriptive and fingerprint clusters in Molecular Operating Environment (MOE). The descriptive clustering is based on the descriptive properties of a small molecule such as a_acid (Number of acidic atoms), a_base (Number of basic atoms), A_hyd (Number of hydrophobic atoms), B_rotN (Number of Rotatable bonds), B_ar (Number of aromatic bonds), B_single (Number of single bonds), B_double (Number of double bonds), B_triple (Number of triple bonds), Lip_don (The number of OH and NH atoms), Lip_acc (The number of O and N atoms), Rings (The number of aromatic/ heterocyclic rings). The descriptive clustering is done using Principle component analysis. (The Principle component analysis is a mathematical procedure used to convert the values of correlating variables into a set of values of uncorrelated variables called principal components). We used the Molecular Operating Environment from the Chemical computing group to perform the cheminformatics analysis.

The fingerprint clustering is based on the FP:MACCS keys. The FP:MACCS keys describe a chemical structure in binary terms. The fingerprint clustering was done with an overlap threshold between 40% and 50% initially and increased to 95% for two clusters since more than half of the fragments appeared in these two clusters. A tanimoto coefficient is used as the metric for similarity purpose. The tanimoto coefficient \( \tau = \frac{N_{AB}}{N_A + N_B - N_{AB}} \) where NA is number of features (ON bits) in A, NB is the number of features (ON bits) in B, and NAB is the number of features (ON bits) common to both A and B for two molecules A and B [126].
2.2.4. Selection of Cluster Representatives

The representatives from each cluster are selected by similarity searches using Sci-finder and Zinc databases. The similarity threshold is maintained at 95% for most of the cluster representatives. Further 75% similarity threshold is implemented for representatives of clusters, which did not reach 95% similarity. Visual inspection to preserve the key functional groups is also ensured during the similarity search. The list of the fragments is sent to Sigma-Aldrich for suggested compounds. The fragments obtained from the similarity search are purchased from commercial sources.

2.2.5. Quality Control (QC) and Plating

NMR spectroscopy is used for assessing the quality of the purchased compounds. Four NMR spectra were acquired for each compound at a concentration of 200μM in 700 μliters 90% H2O, 10% D2O. The buffer contained 10 mM NaCl, 10 mM Na-Phosphate, pH=6, 0.1 mM EDTA. Quality control is run on a 500 MHz spectrometer with an Innova console and an HCN probe and a 50-position sample changer. The four QC spectra included 1d spectra with Watergate (64, 256 or 320 scans, 12 ppm sweep width, 32 K data points), and excitation sculpting (64 scans, 12 ppm sweep width, 32 K data points), respectively for solvent suppression, along with the WaterLogsy (512 scans, 12 ppm sweep width, 32 K data points, 1 ms mixing period) and CPMG (or T2 filter) (64 scans, 12 ppm sweep width, 32 K data points, 400 ms mixing period). Excitation sculpting was used for solvent suppression for the WaterLogsy and T2 filter experiments (the details for Watergate, WaterLOGSY and T2 filter experiments are mentioned in chapter 3). The 1d excitation sculpting experiment used identical parameters, including the same pulse
program as the T2 filter experiment, but with the T2 filter period set to zero. Spectra were converted to a format that was readable in TopSpin 2.0 (Bruker) and processed. Analysis was aided by the use of AMIX software (Bruker). The compounds with spectra that did not fit the expected spectral pattern are eliminated along with the compounds that gave binding signal in the absence of RNA. The fragments that passed the QC are compared to the representatives of the clusters to ensure all the clusters are represented.

2.2.6. Screening of Fragments for Binding to the Bacterial Ribosomal A site RNA

The fragments that passed the QC are used for screening against the bacterial ribosomal A-Site RNA. One-hundred-two compounds were screened in 29 mixtures of 3-4 per tube, containing the same buffer as the QC spectra but with 10μM A-site RNA. A competitor with known binding to the A-site RNA, Gentamycin, was added to a concentration of 30μM, and experiments were repeated with identical parameters for each mixture. Spectra from a hit compound are identified from the following characteristics: a positive signal in the WaterLOGSY spectrum (320 scans, 1.2 ms mixing period) and a reduction in signal in the T2 filter experiment (128 scans, 300 ms T2 filter period) relative to the 1 d experiment with excitation sculpting. We expected that hits that bind specifically to the aminoglycoside binding site show a reduction in positive signal in the WaterLOGSY and some recovery of signal in the T2 filter experiment when Gentamycin is added. However, Gentamycin is a relatively weak affinity ligand compared to competitor ligands commonly utilized in fragment screening against kinases. Ligands that showed small or ambiguous indications of competition with Gentamycin were further tested as described below.
2.2.7. Confirmation of the Fragment Hits

Hit candidates identified from the library screen are tested directly for binding to the A-site RNA. The buffer conditions are 10 mM NaCl, 10 mM Na phosphate pH 6.7, 0.1 mM EDTA in 700 μL of 90% H₂O, 10% D₂O. The concentration of the compounds is between 200-300 μM. WaterLOGSY, T₂ filter, and 1D Watergate and excitation sculpting experiments are acquired on a 700 MHz Varian NMR spectrometer with an HCN cold probe. A long acquisition (number of transients) 1024) 1D Watergate spectrum with a large sweep width (24 ppm) is performed to obtain a spectrum of the RNA Imino resonances.

2.3. Results

2.3.1. RNA Binding Ligand Database

We obtained an RNA binding ligand database of 120 ligands. Each ligand has a reported Kₐ value of less than or equal to 50μM with RNA. A few ligands are included for which the reported Kₐ is not available, if a PDB coordinate set existed showing the ligand bound to an RNA. We incorporated compounds from a variety of sources, including antibiotics, riboswitch effectors, and products of in vitro drug design studies. We intended to limit redundancy in the database while including as much of proven RNA-binding chemical space as possible. Compounds that belonged to similar chemical classes were included if they incorporated distinct chemical moieties that were not present elsewhere in the list. (Refer to [124] for the list with references for each ligand).
2.3.2. Fragment Database

The fragment database is enriched with small molecule fragments that are obtained after the fragmenting and choosing of the representative fragments from the fragment clusters as mentioned in the methods section. We obtained a total of 260 fragments. Many of these fragments are similar in structure. The clustering reduced the fragment redundancy. The clustering based on the fingerprints utilizes 166 MACCS keys that are a set of questions about a chemical structure. Some of the questions are 1) Are there fewer than 3 oxygens? 2) Is there a S-S bond? 3) Is there a ring of size 4? 4) Is at least one F, Cl, Br, or I present? The result is a list of binary values – either true (1) or false (0). This list for a chemical structure is called the MACCS key fingerprint for that structure. The descriptive clustering is based on the physico-chemical properties of a molecule. An example is shown in Figure 2.1 with the molecule on the left and the descriptors on the right.

The clustering based on fingerprint and descriptive properties resulted in 55 descriptive clusters and 53 fingerprint clusters. The cluster number 3 and cluster number 74 of the fingerprint clusters had to be re-clustered because more than half of the fragments showed in these two clusters. The details of the clustering are mentioned in table 2.2. The rationale for re-clustering the fingerprint cluster 3 and 74 is that it could be representing the chemical space that has propensity to RNA and further clustering could result in clusters that densely represent the RNA binding chemical space. We searched for at least one purchasable compound that could represent each cluster, and some subclusters, which had been identified through the above process (methods). A more exhaustive catalogue of clusters and purchased fragments for the library is available in the supplementary information of [124].
of 67 ligands with reported RNA-binding propensity. Data base ligand structures are directly as possible from fragments of compounds with proven RNA-binding propensity. Data base ligand structures were “cleaved” in silico (see Experimental Details). Altogether this process produced 55 descriptive clusters and 53 fingerprint clusters. We wished to represent each cluster in our library, but we noted that a disproportion of these fragments are chemically similar. Therefore, we clustered the fragments, first, based upon chemical descriptors and, second, based upon molecular fingerprints using MOE (from left to right): (a) molecular weight, (b) Chi0v (zero order atomic valence connectivity index) (from CamSoft). These compounds are similar analysis on two publicly available “diversity” (from the Ashgate Drug Index (CamSoft). These compounds are from the Ashgate and NCI diversity set compounds. Pittcon: 327 UPMCSD virtual screening compounds purchased 114 total RNA-binding ligand fragments 260 subclusters of fingerprint cluster 74 19 subclusters of fingerprint cluster 3 71 total RNA-binding ligand fragments 260 total clusters + subclusters 198 compounds purchased 114 fragments passing QC 100

**Table 2.2: Statistics of the clustering and fragment database.**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of descriptive clusters</td>
<td>55</td>
</tr>
<tr>
<td>no. of fingerprint clusters</td>
<td>53</td>
</tr>
<tr>
<td>subclusters of fingerprint cluster 3</td>
<td>71</td>
</tr>
<tr>
<td>subclusters of fingerprint cluster 74</td>
<td>19</td>
</tr>
<tr>
<td>total RNA-binding ligand fragments</td>
<td>260</td>
</tr>
<tr>
<td>total clusters + subclusters</td>
<td>198</td>
</tr>
<tr>
<td>compounds purchased</td>
<td>114</td>
</tr>
<tr>
<td>fragments passing QC</td>
<td>100</td>
</tr>
</tbody>
</table>

**Figure 2.1:** Example of (a) descriptive properties for a molecule in (b).
We searched for at least one purchasable compound that could represent each cluster, and some subclusters, which had been corresponding chemical space regions.

We therefore further clustered these ligand subsets with the aim of achieving a “denser” representation of the corresponding chemical space regions. We judged that these clusters may represent a region of chemical space with particularly favorable RNA-binding propensity.

Altogether, 102 compounds were initially selected, purchased, and for screening the library against RNA. Fragment library compounds are derived from the database of RNA-binding ligands. NMR is used to screen the library for compounds that bind with weak affinity to the RNA target. The complexes of these compounds with the RNA go into the fragment library.

2.3.3. Physico-Chemical Analysis of the Fragments In silico and Quality Control

The cheminformatics tools are used to analyze the physico-chemical properties of the fragment molecules. The fragments have “leadlike” properties. As compared to drugs, they tend to have lower MW, lower hydrophobicity and higher solubility and bind weakly to the target. The fragments have a ClogP (solubility index) value of less than 3 and molecular weight within 75-250. This fragment based approach allowed us to explore the chemical space represented by the bulkier ligands. The quality control as mentioned in the methods section resulted in 102 fragments that are plated as described in methods.

2.3.4. Screening of the Fragment Library with the Ribosomal A-site RNA

We used a 27 nucleotide ribosomal A-site RNA with an internal loop from E.coli 16srRNA (Figure2.3) as the target RNA since it is previously shown to bind to aminoglycoside

Figure2.2: (a) Distribution of number of fragments per cluster derived from fragmenting RNA-binding ligand structures in silico and clustering the resulting fragments according to physico-chemical descriptors. (b) Distribution of number of fragments per cluster derived from fragmenting RNA-binding ligand structures in silico and clustering the resulting fragments according to chemical fingerprints.
antibiotics (refer to chapter 1 for A-site RNA details). The construct shown in the Figure 2.3 is similar to the A-site construct previously used for NMR studies [127]. We chose this target for our first fragment library screen because the same ribosomal site was the target for several of the known RNA-binding ligands that provided the source of our fragment structures.

Figure 2.3a: The secondary structure of the A-site RNA used for screening the fragment library. The G1491 residue is the residue shown to interact with Aminoglycosides (chapter1).

The 102 fragments are screened for A-site binding in 29 mixtures of three to four compounds, each at 200μM concentration, through the acquisition of four 1D NMR experiments on each mixture. Each sample contained 10 μM RNA. The methods section has the details of the buffer components and the NMR experiments. The screening experiments work on the principle of detecting ligand signals, through the transfer of magnetization properties from bound to free signal via chemical exchange. Figure 2.3b shows an example of the screening results for one mixture. Five hit compounds are obtained from the initial screen. The hit fragments are shown in table 2.4. The initial screen is followed by a “competition screen” with the known A-site binding ligand, Gentamycin, present.
Figure 2.3b: Screening data for a sample: The sample contained a mixture including compounds 2 and 4 (table 2.4) and two nonbinding compounds, together with 27 nucleotide A-site RNA. From top to bottom: WaterLOGSY, without and with Gentamycin competitor, 1D spectra with excitation sculpting (T$_2$ filter period of 0 s) and with (0.3 s) T$_2$ filter, without Gentamycin, and repeat of the excitation sculpting 1D with and without T$_2$-filter with Gentamycin present. Competition of 4 for the Gentamycin site is evident from the recovery of signal in the T$_2$ filter experiment, although little competition is observed in the WaterLOGSY. Binding of compound 2 is indicated by signals near 6.9 and 7.3-8.1 ppm in the WaterLOGSY. A nonbinding compound, a furan, gives negative signals between 6.5 and 6.9 ppm.

The competition experiment indicates if any compounds, bind specifically to the known A-site “active-site” (Figure 2.3b). Some of the hit compounds showed reduced evidence of binding as shown in table 2.3. However some hits, such as 2-aminoquinoline (2), which has been reported as a component of an A-site active site binder, did not show reduced evidence of binding upon addition of competitor. This result may indicate that the competitor ligand does not bind tightly enough to fully displace the fragment hit. Alternatively, the fragment may be capable of binding in the same region, or other parts of the RNA, simultaneously with the competitor.
Table 2.3: The fragment hits from initial screen and the results from the competition screen.

<table>
<thead>
<tr>
<th>ID</th>
<th>Structure</th>
<th>Source target</th>
<th>Water Logsy/competition with gentamycin</th>
<th>T₂ Filter/competition with gentamycin</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td><img src="image1.png" alt="Structure 1" /></td>
<td>A-site</td>
<td>+/-competitive</td>
<td>+/- competitive</td>
</tr>
<tr>
<td>2</td>
<td><img src="image2.png" alt="Structure 2" /></td>
<td>A-site</td>
<td>+/-non-competitive</td>
<td>+/- non-competitive</td>
</tr>
<tr>
<td>3</td>
<td><img src="image3.png" alt="Structure 3" /></td>
<td>A-site</td>
<td>Shifted, broad + signal/competitive</td>
<td>+/- competitive</td>
</tr>
<tr>
<td>4</td>
<td><img src="image4.png" alt="Structure 4" /></td>
<td>RNA bulge site</td>
<td>+/-non-competitive</td>
<td>+/- competitive</td>
</tr>
<tr>
<td>5</td>
<td><img src="image5.png" alt="Structure 5" /></td>
<td>Riboswitch</td>
<td>Shifted signal/competitive</td>
<td>+/- competitive</td>
</tr>
</tbody>
</table>

The five hit compounds are verified through follow-up experiments performed at higher sensitivity on a 700 MHz spectrometer with a cold probe and a long acquisition time in order to detect imino signals from the RNA. The latter are diagnostic of RNA base pairing. Chemical shift change and/or line broadening in this region indicate ligand contact with the RNA. Figure 2.4 shows the follow-up screening experiment for compound 4 A-site RNA.
Figure 2.4: Follow up WaterLOGSY and T₂ filter experiments for the compound 4. WaterLOGSY, T₂ filter, are shown with (blue) and without (black) RNA.

Figure 2.5 shows the follow up screen with 10:1 excess of ligand concentration. It is seen from the RNA Imino resonances that Compounds 1, 4, and 5 all interact with the RNA, as indicated by shifts in Imino resonances, notably in the signal for the G1491 residue (near 12 ppm), indicating that all bind near the active site. The spectrum in the second panel from the top, shows no change from the “free RNA” control, indicating that diethyl Nicotinamide does not bind A-site RNA.
2.4. Conclusion

We have learned lessons in conducting RNA targeted screens some of which are as follows. The NMR techniques we used in this screen against the A-site RNA are more sensitive towards weak binding interactions and also require less material (RNA) as compared to other methods which use (just 1d RNA-detected experiments [128]) more

Figure 2.5: RNA Imino resonances from samples containing 10:1 excess of the compounds 4,1,5,3 (bottom to top). The top most panel is “RNA only”.

RNA. Other research groups have used the “Saturation Transfer Difference” NMR experiment to RNA binding ligands [129]. However, our group and others [122] found that the STD is not as sensitive as the WaterLOGSY and the T2 filter experiments for RNA [122].

Our screening method with a rational approach towards identifying ligands with RNA bindability and also amenable for derivatization is designed to rescue the aforementioned RNA binding functionalities in RNA targeted screens [124]. Though screening with “generic” fragment libraries against RNA has resulted in the discovery of some novel ligands previously [128], it is likely to miss such functional moieties responsible for RNA binding through these screens because generic fragment libraries include only drug like molecules according to conventional criteria. Molecules such as those with sugar moieties are not considered good leads though their ability to bind to RNA is established, for example. At this time, clearly definable criteria that can be applied across the range of RNA targets for enhancing the possibility of RNA binding are not uncovered. We chose the fragments to mimic chemical functionalities already proven to bind to RNA in the context of obtaining drug like hits for understanding the properties of ligands that bind to RNA in our context.

This study represents a first step toward the development of an RNA-directed cheminformatics. We manually analyzed the physico-chemical properties of the fragments to obtain ligands that can be specifically used for RNA screening with specific advantages over screening RNA with original or large ligands like Aminoglycosides. The fragments have smaller polar surface area (because of which non specific interactions can be reduced), are amenable for further “growing” to develop specificity and affinity for
specific RNA targets. They are not targeted against a specific target structure in our case (addressing the problem with using SBDD). The fragments that are multiple hits (for different targets) can be used for in silico screening against flexible RNA targets by the probing of target flexibility through multiple conformations observed with multiple hit fragments, fragments will complement in silico screening methods recently designed to incorporate RNA induced fit [129, 130]. However, further iterations of the screening and addition of new fragments to the library will lead to the development of drug leadlike libraries with our proposed fragment based approach. As we hoped that such a development would be realized, a fragment-based strategy has since been used in identifying novel RNA binding fragments for TPP riboswitch and HCV IRES [131].
CHAPTER 3: DETERMINING THE STRUCTURAL ELEMENTS REQUIRED FOR LIGAND DISCRIMINATION BETWEEN ON AND OFF STATE SAM-I RIBOSWITCH CONFORMATIONS

3.1. Introduction

The ongoing discovery of riboswitches in the last one and a half decades, has added a new dimension to our understanding of the role of RNA as a gene regulatory element. Riboswitches are folded RNA elements that sense and bind to certain cellular metabolites and regulate the expression of genes related to the metabolites[132, 133]. Riboswitches are found near the 5’ untranslated region of mRNA in gram positive bacteria, a few plants and in certain fungi [134]. Riboswitch gene control mechanisms include modulation of rho-independent transcription terminator formation, sequestering of ribosome binding sites, formation of ribozymes and control of alternative splicing [74, 135]. Riboswitches appear capable of manipulating the expression of genes related to metabolite biosynthesis without direct interaction with proteins. Moreover, riboswitches have been suggested as potential drug targets ([37, 81, 101, 102, 131, 136-138]), and for a number of potential engineering applications[139-141], highlighting the importance of understanding mechanisms for specificity and recognition of riboswitches by small molecule ligands[139, 142].

The S box system, also called the SAM-I riboswitch was identified by careful sequence data analysis in B.subtilis (Moszer et al., 1995; Kunst et al., 1997) [143]. The SAM-I riboswitch controls the expression of genes involved in sulfur metabolism, synthesis and transport of sulfur-containing amino acids using S-adenosyl methionine (SAM) as the effector molecule[70, 106, 144-146]. The SAM-I riboswitch is one of five classes within
the S-box family. Each class has its own sequence conservation and gene control mechanism [147]. In the SAM-I riboswitch, binding of SAM to the riboswitch causes structural rearrangement of the RNA leading to transcription modulation [148, 149]. The small molecule SAM is the cue for bacterial sulfur metabolism. The remarkable ability of such a small molecule to alter the global structure of the macromolecule (RNA) motivates the search for an in-depth understanding of the regulatory mechanism. Such an understanding in turn will provide insights into the principles of metabolite recognition and discrimination by RNA.

The SAM-1 riboswitch is comprised of two functional domains (Figure 3.1). The first domain called the aptamer binds to SAM at raised cellular SAM concentration. Under these conditions, the second domain, called the expression platform, undergoes structural rearrangement to allow the formation of a stable transcription terminator stem-loop. At low cellular SAM levels, the interplay between the two domains is altered to stabilize the antiterminator (AT). These structural alterations in the expression platform allow read-through transcription [150, 151]. The crystal structures of the SAM-I riboswitch in T.tencongensis upstream from the metF-H2 gene and B.subtilis upstream from the yitJ gene revealed that SAM binds to a pocket formed by four coaxially stacked helices (P1, P2, P3, P4) joined by junctions J1/2, J2/3 and J3/4 (Figure 1) [108]. Mutational analysis and biophysical studies show that a kink-turn mediated pseudoknot formation assists the global folding of the riboswitch cooperatively with magnesium present [152, 153].

Since the riboswitch is believed to undergo conformational changes from transcription ‘On’ state to ‘Off’ state in the presence of SAM, the question arises: **how can a limited set of contacts by a small molecule ligand cause a large-scale rearrangement of**
secondary structure folding in the target macromolecule? The dynamic nature of the expression platform poses a major challenge in understanding the gene control mechanism. In the absence of SAM, the P1 helix is dynamic even in the isolated aptamer [151].

We investigate the structural features of the *T.tencongensis metF*, *B.subtilis yifJ* SAM-I riboswitches that are responsible for the discrimination of the ligand by the riboswitch and also the discrimination of the On and Off states of the riboswitch by the ligand. *T.tencongensis metF*, SAM-I riboswitch (Figure 3.1) *B.subtilis yifJ* SAM-I riboswitch (Figure 3.2) are shown. As is clear from these Figures, the base pairing patterns in the P1 helix and the AT helix are similar to each other, as is required by the strand switching mechanism.

![Figure 3.1: The secondary structure of the *T.tencongensis metF* SAM-I riboswitch. The residues in green are conserved. The similarity in base pairing between the P1 helix and the AT is shown in the boxed region. (Adapted from [58]).](image)
This observation, led us to the design of RNA constructs to 1) mimic the On state of the *T.tencongensis* SAM-1 riboswitch (On constructs) (Figure 3.3), 2) mimic aspects of an Off state tertiary structure, within the context of a putative On state secondary structure in the *B.subtilis yitJ* SAM-I riboswitch (Hybrid constructs). Our preliminary SAM binding results with the On constructs and the hybrid constructs indicate that the hybrid constructs (Figure 3.4) of the *B.subtilis yitJ* SAM-I riboswitch bind the SAM with moderate affinity. We examine the basis for discrimination by SAM between On and Off state riboswitch RNAs, in terms of secondary and tertiary structure using the *B.subtilis yitJ* SAM-I riboswitch as a model system.

Figure 3.2: The Secondary structure and sequence of *B.subtilis yitJ* SAM-I riboswitch. The residues of the junction regions J1/2, J3/4 and J1/4 are colored in red, pink and blue respectively. The aptamer, AT and T are decision-making regions that represent the aptamer, antiterminator and the terminator sequences. The P1-P4 are helices of the SAM-I riboswitch.
While ligand affinity is much reduced compared to that for the aptamer, binding to the hybrid constructs shows a similar pattern of sensitivity to mutations and ionic conditions as reported for the former. Thus, the mutant binding data are consistent with the conclusion that the hybrid RNAs mimic aspects of aptamer tertiary structure, which mediate SAM binding. Consistent with previous indications that a minimal length P1 helix is necessary for SAM binding, continuous stacking between P1 and AT helical regions is required for SAM binding to hybrids. We discuss the implications for models of conformational switching mechanisms for the SAM-I riboswitch and other riboswitch systems.

Figure 3.3: T.tencongensis On constructs GGU_P1 (left, the residues in red are the altered residues to mimic the P1 helix), GGAG_AT_U4 (right).
Figure 3.4: *B. subtilis yif* SAM-I riboswitch AT and AAT hybrid constructs used for SAM binding studies. Helices P2, P3, P4 and the junctions regions J1/2, J3/4, J1/4 are represented with sticks as they remain consistent in all the hybrid constructs.

3.2. Methods and Materials

3.2.1. Materials

Synthetic DNA oligos for generating DNA templates by PCR were purchased from Integrated DNA technologies (IDT), Inc (USA). Equilibrium dialysis (Dispo equilibrium dialyzer 5000 MWCO) chambers were purchased from Harvard apparatus, Massachusetts, USA. The S-adenosyl-L-methionine-methyl-\(^{3}\text{H}\), \(^{32}\text{P}\) labeled ATP for 5’RNA end labeling are purchased from Perkin Elmer, California, USA. The NTPs for in vitro transcription were purchased from Sigma Aldrich.
3.2.2. Methods

3.2.2.1. DNA Template Synthesis

The synthetic oligo nucleotides purchased from IDT were purified by denaturing PAGE on 12% acrylamide gel, electro-eluted and recovered by ethanol precipitation. Oligo nucleotide concentration was determined using U.V absorbance at 260nM.

3.2.2.1.1. Cassette Polymerase Chain Reaction for *B.subtilis yitJ* SAM-I Riboswitch with Hammerhead Ribozyme

The long templates for transcription reactions were synthesized by creating a cassette with 2 rounds of PCR reactions. First round of PCR reactions were done with over-lapping oligo nucleotides [154]. Details of the sequences used for each round and the design for the hammerhead ribozyme cleavage are included in table 3.1.

The T7 promoter and hammerhead ribozyme coding sequences are incorporated in the second round of PCR. A schematic for the cassette PCR is shown in the Figure 3.5. Similar DNA templates for all the hammerhead ribozyme integrated hybrid constructs are generated for SAM binding experiments using cassette PCR strategy.

PCR reactions are done in 100μL with 2mM MgCl₂, 10PM DNA template strands, 1μM primers, 200μM dNTPs and taq polymerase. The PCR cycle includes a 94°C denaturation step, 58-62°C annealing step (based on the primer melting temperatures), 72°C extension step. 31 cycles for first round PCR were done. DNA from the first round PCR is used as template for the second round PCR along with the appropriate primers (modified primers are used to generate the template strand with 5’ modified nucleotides to ensure homogenous 3’ ends of RNA) and the hammerhead ribozyme coding bottom strand
Figure 3.5: Schematic representation of cassette PCR strategy. The first round of PCR cycles generates the riboswitch coding sequence represented in black color. The second round of PCR cycles generates riboswitch (black), hammerhead ribozyme (cyan) and T7 promoter coding regions (green). To produce templates for RNA transcription, the following DNA oligonucleotide sequences (Overlapping templates) were added to a PCR reaction for the first stage of the cassette strategy.

The second round PCR was done in similar conditions as the first round. The amplified DNA on a 2% agarose gel is shown in Figure 3.6.

Table 3.1: List of yitJ SAM-I riboswitch hybrid constructs and their primers used to generate the templates.

<table>
<thead>
<tr>
<th>Riboswitch construct</th>
<th>Forward primer</th>
<th>Reverse primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>2P1_11AT</td>
<td>5’ATC AAG AGA AGC AGA</td>
<td>5’m10mA AAG AAGAC</td>
</tr>
<tr>
<td></td>
<td>GGG ACT G3’</td>
<td>TTT GTC ACT GAT TTT3’</td>
</tr>
<tr>
<td>3P1_10AT</td>
<td>5’ ATC AAG AGA AGC</td>
<td>5’m10mA AAG AAGAC</td>
</tr>
<tr>
<td></td>
<td>AGA GGG ACT CG 3’</td>
<td>TTT GTC ACT GAT TTT3’</td>
</tr>
<tr>
<td>4P1_9AT</td>
<td>5’TAT CAA GAG AAG CAG</td>
<td>5’m10mA AAG AAGAC</td>
</tr>
<tr>
<td></td>
<td>AGG GAC TG3’</td>
<td>TTT GTC ACT GAT TTT5’</td>
</tr>
<tr>
<td>6P1_7AT</td>
<td>5’CTT ATC AAG AGA AGC</td>
<td>5’a1mA mG AAGAC TTT</td>
</tr>
<tr>
<td></td>
<td>AGA GGG ACT CG 3’</td>
<td>GTC AGT GAT TTT3’</td>
</tr>
<tr>
<td>8P1_5AT</td>
<td>5’TTC TTA TCA AGA GAA</td>
<td>mGm AAGAC TTT GTC</td>
</tr>
<tr>
<td></td>
<td>GCA GAG GGA3’</td>
<td>AGT GAT TTT GT3’</td>
</tr>
<tr>
<td>6P1_10AT</td>
<td>5’CTT ATC AAG AGA AGC</td>
<td>mAm AAGAC TTT GTC</td>
</tr>
<tr>
<td></td>
<td>AGA GGG ACT CG 3’</td>
<td>TTT GTC ACT GAT TTT5’</td>
</tr>
<tr>
<td>8P1_10AT</td>
<td>5’TTC TTA TCA AGA GAA</td>
<td>5’m10mA AAG AAGAC</td>
</tr>
<tr>
<td></td>
<td>GCA GAG GGA3’</td>
<td>TTT GTC ACT GAT TTT3’</td>
</tr>
<tr>
<td>3P1_9AT</td>
<td>5’ATC AAG AGA AGC</td>
<td>5’m10mA AAG AAGAC</td>
</tr>
<tr>
<td></td>
<td>AGA GGG ACT CG 3’</td>
<td>TTT GTC ACT GAT TTT5’</td>
</tr>
<tr>
<td>3P1</td>
<td>5’ ATC AAG AGA AGC</td>
<td>5’ATC TTC CAA GCT GTT</td>
</tr>
<tr>
<td></td>
<td>AGA GGG ACT CG 3’</td>
<td>CG</td>
</tr>
<tr>
<td>3P1_10AT_G11</td>
<td>5’ATC AAC AGA AGA AGA</td>
<td>5’m10mA AAG AAGAC</td>
</tr>
<tr>
<td></td>
<td>GGG ACT G3’</td>
<td>TTT GTC ACT GAT TTT3’</td>
</tr>
<tr>
<td>3P1_10AT_G5CC</td>
<td>5’GCC ATC AAG AGA AGC</td>
<td>5’m10mA AAG AAGAC</td>
</tr>
<tr>
<td></td>
<td>AGA GGG ACT CG 3’</td>
<td>TTT GTC ACT GAT TTT3’</td>
</tr>
</tbody>
</table>
5’ GGT TCT TAT CAA GAG AAG CAG AGG GAC TGG CCC GAC GAA GCT TCA GCA ACC GGT GTA ATG GCG ATC AGC CAT GAC CAA GGT GCT AAA TCC AGC AAG CTC GAA 3’

5’ AAA GTC CTC TTA AGA AGA AGA CT T TGT CAG TGA TTT TGT CTC TTT TTA TCT TCC AAG CTG TTC GAG CT T GCT GGA TTT AGC ACC TTG GTC ATG GCT GAT CGC CAT3’

3.2.2.1.2. Cloning for Generating DNA Templates for the In vitro Transcriptions of the *T.tencongensis* SAM-I Riboswitch Constructs

The *T.tencongensis* aptamer and the “ON conformers” are synthesized by in vitro transcriptions run on DNA templates generated by plasmid PCR reactions.

Figure 3.6: A 2% ethidium bromide stained agarose gel with 50 base pair ladder (lane 1), the *yitjP1_AT* lanes (2,3) are the DNA products from the first round of PCR, the *yitjP1_AT_RIBOZYME* lanes (4,5,6,7) are the DNA from second round of PCR.

The DNA insert coding for the riboswitch construct is ligated into a restricted plasmid with a HDV ribozyme coding sequence and T7 promoter coding sequence. The PCR to generate the insert was done according to methods mentioned in section 3.2.2.1.1. The
plasmid is linearized according to the EcoR1 and NcoI restriction enzyme (New England Biolabs) requirements (table 3.2).

**Table 3.2: Reagents and materials for restriction reactions**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Plasmid (μL)</th>
<th>Insert DNA from PCR product (μL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂O</td>
<td>10</td>
<td>13</td>
</tr>
<tr>
<td>Buffer (NEB 10x)</td>
<td>2.0</td>
<td>2.0</td>
</tr>
<tr>
<td>EcoR1</td>
<td>1.0</td>
<td>1.0</td>
</tr>
<tr>
<td>DNA</td>
<td>8.0 (1pM)</td>
<td>6.0 (1pM)</td>
</tr>
<tr>
<td>CLP+ 90 min incubate 37°C</td>
<td>1.0</td>
<td>0.0</td>
</tr>
<tr>
<td>NCOI+ incubate 90 min at 37°C</td>
<td>1.0</td>
<td>1.0</td>
</tr>
</tbody>
</table>

The ligation reaction with the gel purified restricted insert and the plasmid was done at was done using T4 DNA ligase (Promega) at 16°C for 12 hours (table 3.3).

**Table 3.3: Ligation Reactions for Restricted Plasmid and Insert.**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>1:1 Plasmid (μL)</th>
<th>1:3 Plasmid (μL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Plasmid</td>
<td>10</td>
<td>4</td>
</tr>
<tr>
<td>Insert</td>
<td>10</td>
<td>10</td>
</tr>
<tr>
<td>Buffer</td>
<td>2.5</td>
<td>2.5</td>
</tr>
<tr>
<td>H₂O</td>
<td>1.5</td>
<td>7.5</td>
</tr>
<tr>
<td>Enzyme</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>

Transformation of the plasmid is by incubating 50μl of competent cells with 2-10ng of the plasmid DNA for 30 minutes with gentle mixing followed by heat shock for 45 seconds at 42°C. The reaction is placed on ice for 10 minutes. The transformed cells are
grown on LB agar plates with Ampicillin. The clones have been confirmed by colony PCR reactions and by sequencing.

The templates for the in vitro transcription of the *T.ten congensis* SAM-I riboswitch constructs are obtained by colony PCR reactions with appropriate HDV and T7 primers. The primers are T7 promoter (5’, GCGGCGGAATTCTAATACGACTCACTATAG, 3’) and the HDV ribozyme in the vector (5’, GAGGTCCCATTCATTCGCCATGCCGAAGCATGTTG, 3’).

### 3.2.2.2. In vitro Transcriptions

The in vitro transcriptions for the hammerhead ribozyme integrated *B.subtilis* and the HDV ribozyme integrated *T.ten congensis* riboswitch constructs are done in 40 mM MgCl2, 8.0 mM NTPs from Sigma Aldrich, 80 mg/mL 8000 MW PEG, and 100μL PCR template in 250 μL reactions. The hammerhead ribozyme cleavage is attained after 4 hours of transcription at 37°C without any excess magnesium. HDV ribozyme cleavage required 20mM excess of magnesium at 60°C for 20 minutes. The RNA is purified on a 12% acrylamide denaturing gel containing 7M Urea. The appropriate gel slices with the riboswitch are electroeluted, ethanol precipitated overnight and dialyzed in high salt and low salt buffers followed by desalting in PD10 columns. The RNA is concentrated by lyophilizing or using 3000 MWCO Amicon filters. The RNA was refolded by heating to 95°C and snap cooled on ice and supplemented with 2mM MgCl2 and used for binding experiments. Each RNA was run on ethidium bromide stained 12% acrylamide analytical gel for size, concentration and integrity verification.
3.3.2.3. Equilibrium Dialysis

Single point SAM binding assays are done with 100nM SAM and 3μM RNA according to established protocols [106, 155]. 100nM SAM (10.0Cl/mmol, 12000 counts per minute) in a total volume of 30uL in equilibrium dialysis buffer (500mM KCl, 50mM Tris HCL pH 7.5, 20mM MgCl2) was added to chamber A of Dispo equilibrium dialyzer. The RNA to be tested for binding is heated to 95°C for 5 minutes and snap cooled on ice for 10 minutes before the binding experiment. 3μM RNA to be tested for binding in a total volume of 30μL in equilibrium dialysis buffer is added to chamber B and equilibrated at room temperature for 12 hours with shaking. Aliquots from each chamber are collected and counted separately in the scintillation counter. The ratio of number of counts in chamber B to chamber A is used as a measure of SAM binding. Averaging of at least three replicates with standard deviation is used as the B/A ratio for each RNA.

3.2.2.4. SAM Titration (Scatchard Analysis)

Affinity of SAM to B.subtilis aptamer is calculated from the data obtained from scatchard analysis using equilibrium dialysis to quantitate SAM binding as a function of concentration. 200nm B.subtilis yitJ aptamer/hybrid RNA was added to chamber B of the Dispo equilibrium dialyzer and various concentrations of tritiated SAM (10.0Cl/mmol, 12000 counts per minute) ranging from 10nM to 3 μM is added in chamber A along with equilibrium dialysis buffer (500mM KCl, 50mM Tris HCL pH 7.5, 20mM MgCl2). 30uL was the total volume in each chamber. Degradation of SAM was taken into account by eliminating 10% of the counts in each chamber. Affinity of SAM to the aptamer/hybrid is calculated from the scatchard plot, plotted using SAM bound (Lb), free SAM (Lf), total SAM
(Lt) were all calculated from the number counts in chamber A and chamber B [106, 155]. The outlier points obtained using dialysis chambers due to leaks in the membrane and air bubbles are eliminated.

Figure 3.7: Secondary structure of the hammerhead ribozyme integrated *B. subtilis* SAM-I riboswitch hybrid construct. The arrow indicates the point of ribozyme cleavage.

3.2.2.5. Inline Probing Assay

Inline probing reactions with the 3P1_10AT, 3P1_9AT and 3P1_10AT_5’GCC hybrid constructs. The RNA required was generated as mentioned in the in vitro transcriptions above. The dephosphorylation and 5’end $^{32}$P labelling, inline probing reactions in the presence of 100µM SAM are done according to the methods mentioned in [106]. RNA dephosphorylation reactions are done with 10pM RNA in the 20µL reactions with 1unit/µL Shrimp alkaline phosphatase enzyme and 10x reaction buffer and incubated at 37°C for 30 minutes. The dephosphorylated RNA is purified by phenol chloroform extraction, ethanol precipitation. 5’ end labeling of the dephosphorylated RNA is done in
the presence of 4µL of 5x phosphorylation buffer, 1uL of 10µM dephosphorylated RNA, 6µl of (γ-32P) ATP (adenosine triphosphate) (6000mCi/mmole), 2.5µL of 10U/µL T4 polynucleotide kinase in a total volume of 10µL and incubated at 37° C for 30 minutes. The labeled RNA is purified by ethanol precipitation. The inline probing reactions are done by incubating 1µL labeled RNA in 5µL of inline reaction buffer and 100µM metabolite at room temperature for 40 hours. Denaturing 10% PAGE and its analysis using molecular dynamics phosphor imager was done to analyze the RNA cleavage patterns.

3.2.2.6. BPP Calculation

The base pair probability was calculated at 37 °C using the RNAfold program in the Vienna RNA package. The probabilities of base pairs in the proposed AAT and AT models are plotted according to the position as annotated in the Figure 2b. Two versions of sequences are used in the calculation. The one without terminator (T) includes the sequence to form either AAT or AT without the 3’ strand of the terminator. The other one is the full-length sequence of the riboswitch, in which the terminator can completely form.

3.2.2.7. Nuclear Magnetic Resonance (NMR) Spectroscopy Experiments

1d, WaterLOGSY and T₂ filter experiments are used to detect SAM and small molecule binding with the different RNA constructs. The binding experiments work on the principles involved in the transfer of magnetization properties from the ligand to RNA[124]. The NMR samples were made with 200-250 µM ligand, 10-20µM RNA in a total
of 500μL with 10mM sodium phosphate, 10mM sodium chloride and, 0.2mM EDTA with or without magnesium. Data was collected on 500Mhz Varian spectrometer.

3.3. Results

3.3.1. SAM Binding with the T.tencongensis SAM-I Riboswitch Aptamer and On Conformers as Seen from 1d Proton NMR

The T.tencongensis metF riboswitch aptamer (henceforth referred to as “T. tencongensis aptamer”) SAM binding is monitored using NMR 1d experiments. The ligand peaks experience changes in the chemical environment upon interacting with the RNA in ligand detected NMR spectroscopy. Such changes are readily observed in WaterLOGSY and T₂ Filter experiments. The WaterLOGSY experiment is a highly sensitive NMR technique that has applications in detecting small molecule interactions with macromolecules. The WaterLOGSY is based on the transfer of the magnetization to the small molecule (SAM) from the macromolecule (RNA) through water molecules [156]. The transfer that takes place is from water bound to the macromolecule and the ligand. If the ligand is bound, then when it exchanges protons with water that leads to NOEs with water, and these show up with a positive sign because the ligand is bound to the RNA and tumbling slowly. The bound water molecules present at the interface of ligand-RNA interaction contribute to the high sensitivity of the WaterLOGSY experiment. The T₂ filter experiment works by monitoring the T₂ relaxation time of the small molecule (SAM) [157]. The small molecule “tumbles” faster than a large molecule (RNA) but slowly when associated with a large molecule. The intensity of the ligand peak decreases when it interacts with the RNA as compared to the peak intensity of the free ligand. In this case, the intensity of the ligand
peak in a conventional one-dimensional experiment may also decrease and the peak may broaden due to exchange between free and bound forms. In RNA detected NMR spectroscopy, the RNA peaks show change in the position or intensity or type of peaks when the RNA interacts with a small molecule ligand. 1D proton Watergate experiments are used to suppress the large water signal for RNA detected NMR. We use these methods to detect the SAM binding to the *T.tencongensis* SAM-I riboswitch constructs.

### 3.3.2 The Aptamer Binds to SAM as Seen from 1d Proton NMR and Equilibrium Dialysis

![Figure 3.9](image)

**Figure 3.9:** WaterLOGSY experiment for SAM in the presence and absence of *T.tencongensis* SAM-I riboswitch aptamer (10 μM aat). The spectrum is split into two regions for scaling. The peaks in blue are the ligand peaks that increase in intensity upon binding to the RNA.
The aptamer construct of the *T. tencongensis* SAM-I riboswitch shows binding to SAM in the WaterLOGSY experiment as shown in Figure 3.9.

Equilibrium dialysis experiments are done to validate the SAM binding to the *T. tencongensis* SAM-I riboswitch aptamer. The TPP *thiC* riboswitch was used as a negative control. As expected, the SAM-I aptamer showed high affinity binding with a b/a ratio of ~8 (Well within the range reported in literature using equilibrium dialysis). The TPP aptamer did not show any binding.

![Figure 3.10: Equilibrium dialysis data for *T. tencongensis* aptamer, TPP aptamer. The SAM in the SAM only and the TPP aptamer dialysis apparatus equilibrates equally between the two chambers shown by a b/a ratio of ~1.](image)

### 3.3.3. The SAM Binding to *T. tencongensis* SAM-I Riboswitch ON Constructs as Observed by NMR and Equilibrium Dialysis

The *T. tencongensis* On construct GGAA_AT_U4 (Figure 3.3) shows SAM binding in NMR experiments. The WaterLOGSY and the T₂ filter experiments are optimized with temperatures from 10-45°C (data not shown) and concentrations of RNA from 10-25μM and 20mM magnesium.

The On construct GGAG_AT_U4 shows clear binding in both the WaterLOGSY and the T₂ filter experiments. However the equilibrium dialysis data (Figure 3.15) at this time did
not show significant binding implying that the NMR techniques picked up some weak interactions between the ligand and the SAM-I riboswitch constructs. The WaterLOGSY technique, which is highly sensitive to weak binding interactions, did not show any binding to the TPP thiC aptamer.

The SAM binding data from the T.tencongensis aptamer and the On conformers indicates that the SAM makes some interactions with the On conformers that can be detected using NMR but not with equilibrium dialysis.

Figure 3.13: $T_2$ filter SAM binding experiment with the GGAG_AT_U4 construct in the presence of magnesium at 25 °C. We used the small molecule dCTP as a positive control. The peaks highlighted in grey are from SAM and the peaks in black box are from dCTP.
Figure 3.14 WaterLOGSY experiment for GGAG_AT_U4 construct at 25°C. The peaks in the shaded region represent the SAM peaks. In the absence of Mg, the SAM peaks near 8.1ppm show an increase in the intensity. With the addition of magnesium, the peaks at 2.0ppm show an increase in the intensity. The peaks in shaded region are from SAM and from dCTP (positive control) shown in black box.

Figure 3.15: Equilibrium dialysis data for the *T.tencongensis* On constructs.

Equilibrium dialysis is a powerful technique in quantifying the small molecule binding to macromolecules. It can be used to detect weak binding but is not sensitive towards long range or short-lived interactions between the small molecule and the macromolecule. This preliminary data is useful for us to design more SAM-I riboswitch constructs that are useful in probing the structural elements of the riboswitch necessary for SAM detection and its binding to the riboswitch. The following sections focus on the *B.subtilis yitf* SAM-I
riboswitch that is similar to the *T.tencongensis* SAM-I riboswitch in sequence and secondary structure (Figures 3.1,3.2)

### 3.3.4. *B. subtilis* yitJ SAM-I Riboswitch SAM Binding

#### 3.3.4.1. BPP Calculations and RNA Secondary Structure Folding Data

In a transcriptional riboswitch, the expression platform determines whether transcription is terminated. A competition between the formation of a terminator or antiterminator (AT) hairpin decides the outcome. For the SAM-I riboswitch, the formation of the AT involves a competition with the P1 helix within the aptamer. Since the riboswitch region encompassing the competition between these two helices is transcribed before the terminator hairpin is, the AT/P1 competition is likely to bias the final outcome of riboswitch-regulated transcription.

Until now, most experimental studies of the dynamics of riboswitch expression platforms have utilized a two-state model for RNA secondary structure. The chemical and enzymatic probes used, however, do not provide quantitative measures of riboswitch conformer populations. Predicted secondary structures, base pairing probability predictions at minimal free energy and comparative sequence analysis can provide a basis for interpreting riboswitch-folding data. While predictions of a single lowest energy structure (the so-called “MFE”) are often used for this purpose, the secondary structure of any given RNA in solution is determined by the distribution of the RNA's secondary structures at thermodynamic equilibrium [158, 159].
Here we used nucleic acid base pairing probability calculations, based upon a partition function, to simulate the competition between the P1 and the AT of the \textit{B. subtilis yitJ} SAM-I riboswitch. Extensive qualitative data regarding the folding of this riboswitch has been reported. The base pairing probability simulations (Figure 3.16) predict that for a transcript length at which the AT helix has been fully transcribed both the P1 (green) and AT (red) helices are populated. The sum of the probability of forming base pairs at positions 4,5,6,7 represented by (black) is almost equivalent to 1. Moreover, the predicted BPP for P1 helix formation declines and that for AT helix formation increases for each position moving down the sequence as presented in Figure 2a. This result predicts that the RNA has a propensity to form thermodynamic intermediates that can allow the formation of base pairs in the P1 and also the AT helix simultaneously.

![Figure 3.16](image)

**Figure 3.16: Base pairing probability calculation for \textit{B.subtilis yitJ} SAM-I riboswitch (selected sequence shown in the right). Positions of the base pairs are numbered from 1 to 12.**

This prediction raises the question as to whether SAM would bind to these “intermediate” conformers. Addressing this question can help delineate the determinants of SAM binding.
specificity for the aptamer structure, since discrimination between folds should be crucial for the riboswitch mechanism.

3.3.4.2. A SAM-I Riboswitch Hybrid Binds to SAM with Affinity in the Micromolar Range

The sulfur anionic moiety of the SAM interacts electrostatically with the minor groove of the first two base pairs of the P1 helix [108]. We constructed a model RNA molecule that is constrained to form a hybrid structure containing the first three base pairs of the P1 helix, and ten base pairs of an AT helix (3P1_10AT). This construct resembles structures predicted as minor conformers by the calculations in Figure 3.16. Residues in the T are truncated to constrain the formation of the AT (Figure 3.4).

The dissociation constant for the SAM-I riboswitch aptamer and the hybrid (3P1_10AT) of SAM were determined using scatchard analysis from equilibrium dialysis (Figure 3.18) [149]. The dissociation constant for the aptamer measured in this way is 32 nM. This value is within the range of what has been previously reported (table 3.3). Scatchard Analysis yielded $K_d$ for SAM binding to the hybrid 3p1_10AT of $\sim$790nM. As a negative control, SAM binding was tested by equilibrium dialysis for ribosomal A-site RNA and a putative Mycobacterium tuberculosis thiC riboswitch aptamer RNA control (data not shown). No potential SAM binding is detected with the A-site RNA or the thiC riboswitch.

The $K_d$ estimates for the aptamer and 3P1_10AT calculated using the single point equilibrium dialysis data obtained with 3µM RNA and 100nM SAM agree with the value captured by scatchard analysis to +/-15%. For subsequent measurements on other RNA
constructs we used single point measurements at ranges of RNA and SAM concentrations that yielded the highest sensitivity data.

**Table 3.3: Reported $K_d$ values for different SAM-I riboswitch constructs determined using different methods.**

<table>
<thead>
<tr>
<th>Riboswitch name</th>
<th>Length of the riboswitch used to measure $K_d$ (nucleotides)</th>
<th>Method used to determine $K_d$</th>
<th>$K_d$ value (reference)</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. subtilis yitj</em> SAM-I</td>
<td>251</td>
<td>Inline probing using RNAse T1 cleaving at G residue</td>
<td>~200nM [155]</td>
</tr>
<tr>
<td><em>B. subtilis yitj</em> SAM-I</td>
<td>124</td>
<td>Inline probing using RNAse T1 cleaving at G residue</td>
<td>~4nM [155] 5nM [160] 30nM [160]</td>
</tr>
<tr>
<td><em>B. subtilis yitj</em> SAM-I</td>
<td>124</td>
<td>Scatchard analysis with equilibrium dialysis assay</td>
<td>10nM [155]</td>
</tr>
<tr>
<td><em>B. subtilis yitj</em> SAM-I</td>
<td>124</td>
<td>Fluorescence</td>
<td>5.3nM [161]</td>
</tr>
<tr>
<td><em>B. subtilis yitj</em> SAM-I</td>
<td>124</td>
<td>SHAPE probing</td>
<td>19nM [162]</td>
</tr>
<tr>
<td><em>B. subtilis yitj</em> SAM-I</td>
<td>165</td>
<td>Filter binding assay</td>
<td>20nM [163]</td>
</tr>
<tr>
<td><em>T. tencongenis</em> metF aptamer</td>
<td>111</td>
<td>Fluorescence</td>
<td>36nM [164]</td>
</tr>
<tr>
<td><em>T. tencongenis</em> metF SAM-I</td>
<td>95</td>
<td>Isothermal titration calorimetry</td>
<td>130nM [165] 1.35nM [147]</td>
</tr>
</tbody>
</table>
Figure 3.17: Scatchard analysis for a) B. subtilis yitJ SAM-I riboswitch aptamer. The sequence used for this analysis is in Figure 3.2 that includes residues 1 to 119. b) B. subtilis yitJ SAM-I riboswitch 3P1_10AT hybrid shown in Figure 3. The variable r is the ratio of bound ligand to total RNA concentration. The variable L_F represents free ligand [106].

3.3.4.3. Sensitivity of SAM Binding to Ions and Mutations for Hybrid Constructs Parallels That for the Aptamer

Metal ions play an important role in secondary structure folding of RNA [166]. [167]. In particular, Mg has been shown to be required for optimal binding of SAM to the SAM-I riboswitch and for aptamer induced folding [152]. X-ray structures of the SAM bound SAM-I riboswitch, and MD simulations derived from the X-ray coordinates, indicate the critical role played by Mg ions in facilitating ligand interactions with P1, P3 and J1/2 region [108, 152, 164, 168, 169]. We reasoned that if tertiary folding and the mode of SAM binding in the 3P1_10AT mimics that of the full-length aptamer, then magnesium ions will be required to obtain optimal SAM binding to 3P1_10AT. Measurements in the presence and absence of magnesium (Figure 3.18a) clearly indicate the necessity of magnesium for optimal SAM binding. The B/A ratio decreased by more than two fold (as compared to
binding with the snap cooled RNA in presence of magnesium) when 1) Freshly synthesized RNA is not snap cooled. In this case barely any binding was detected (Figure 3.18a), 2) Freshly synthesized RNA snap cooled in the absence of magnesium bound weakly in a detectable range of B/A ratio above 1.5, 3) RNA samples that were pre-hybridized with Mg and stored at -20°C were found to lose SAM-binding activity during storage, but SAM binding was restored by repeating the heat/snap cooling process (data not shown).

![Graph and diagram](image)

**Figure 3.18** a) Ratio of Number of counts in chamber b to chamber a in equilibrium dialysis for *B. subtilis* yitJ SAM-I riboswitch aptamer and 3p1_10AT hybrid in the presence and absence of magnesium dependent folding. b) *B. subtilis* yitJ SAM-I riboswitch 3P1_10AT pseudoknot mutants. c) Equilibrium dialysis data for Aptamer (AAT) (blue), Pseudoknot mutant (pink), pseudoknot compensatory mutant (green), 3P1_10AT (purple) and 3P1_10AT_G11 mutant (cyan). Equilibration of SAM in the two chambers is shown with SAM only (orange). The error bars represent the standard deviation for three replicates for experiment.
It was reported that mutation of the two guanine residues in the apical loop of the P2 helix in the aptamer (boxed nucleotides in the Figure 3.18b) decreased the SAM binding, while compensatory mutation induced by mutating the J3/4 region restored the affinity to SAM [161, 170]. This pattern reflects the formation of a kink-turn-stabilized pseudoknot interaction between the P2 loop and J3/4 [152]. We introduced the same mutations in the SAM-I riboswitch hybrid 3P1_10AT pseudoknot region (3P1_10AT_P2 mutant). In the equilibrium dialysis assay, the 3P1_10AT_P2 mutant decreased the ratio of counts in the chamber containing to RNA to that containing SAM only to less than half (Figure 3.18c) as compared to 3P1_10AT. Compensatory mutation in the J3/4 designed to restore pseudoknot base pairing, restores the affinity to SAM. These trends match those reported in SAM binding studies to the wild type yitJ SAM-I riboswitch aptamer [161, 170].

The G11 residue of the aptamer in junction J1/2 with G11.C44–G58 triple makes contact with the methionine moiety of SAM through hydrogen bonds. The carboxylate group of SAM is recognized by the G11 via the N1 and N2 Watson-Crick face [108, 168]. We hypothesized that mutating the G11 residue will decrease SAM binding to the 3P1_10AT hybrid, as reported within the context of the aptamer alone [171]. As expected mutating the G11 residue decreased the SAM binding drastically in the 3P1_10AT construct to undetectable levels (Figure 3.18c).

3.3.4.4. SAM Binding Requires Stabilization of the P1 Helix by Coaxial Stacking with an AT Helix in Hybrid Constructs

The P1 helix minor groove forms one side of the binding pocket for SAM as shown in the SAM-I riboswitch crystal structures [108, 168]. The sulfur anionic moiety of the ligand
interacts electrostatically with the minor groove of the first two base pairs of the P1 helix[108, 168]. Mutations of these P1 helix residues [106, 172], disruption of P1 helix base pairs, or shortening of the P1 helix [172] dramatically reduces SAM binding to the aptamer. To test the role of the AT in SAM binding in the hybrid, we tested SAM binding with ‘3P1_only’ with RNA concentration of 8µM and ligand concentration 100nM, (Figure 5a). The 3P1_only construct that lacks the complete AT and consists of only 3 base pairs in the P1 helix yielded a B/A ratio of ~1.2 in the equilibrium dialysis assay, indicative of minimal SAM binding. This minimal SAM binding observed with 3P1 only indicates that the AT segment contributes to SAM binding in the hybrid construct. We reasoned that the AT helix might be forming a coaxial stack with three P1 base pairs, thus stabilizing the P1 helix formation that is necessary for SAM binding [106, 107, 172]. To test this possibility, we deleted a residue in the 3’ end of the AT helix (3P1_9AT), thus preventing the possibility of optimal coaxial stacking. As expected the 3P1_9AT construct shows minimal binding with SAM with a B/A ratio of 1.16 (Figure 3.19). We utilized in-line probing to obtain a more direct indication of P1 helix formation in the 3P1_9AT, 3P1_10AT and 3P1_10AT_5’GCC constructs (Figure 3.2a, 3.2b). The 3P1_10AT_5’GCC construct is shown in Figure 3.21a. It is designed to create instability in the 3P1_10AT hybrid in the P1-AT stacking region. Unlike chemical modification methods, which rely on primer extension to detect chemical reactivity, this technique enabled us to work with the hybrid constructs with fixed 3’ ends. As expected, the addition of SAM leads to changes in the pattern of cleavage for 3P1_10AT (Figure 3.21b).
Figure 3.19: The ratio of number counts in chamber b to chamber a in equilibrium dialysis assay for B. subtilis yitJ aptamer, and other hybrid constructs. SAM binding observed for the 3P1_10AT hybrid is reduced dramatically when the possibility for optimal coaxial stacking with the AT helix is disrupted. The error bars represent standard deviation for at least three replicates of each experiment.

These include protection of J3/4 residues involved in pseudoknot formation and enhanced cleavage of a G residue closing the P4 helix. The conserved three-nucleotide motif, GAU, which participates in the putative 3P1 helix, is protected in the presence of SAM, while enhanced cleavage is observed in the adjoining J3/4. This pattern parallels that reported for SAM binding to aptamer or full length yitJ riboswitch constructs[106]. A similar but more modest pattern of SAM-induced changes in cleavage levels is observed for 3P1_9AT, consistent with a lower level of SAM binding as compared to 3P1_10AT_5’GCC and 3P1_10AT.

3.3.4.5. SAM Binding to the Hybrids Increases Marginally as the Length of the P1 Helix Increases

Previous reports show that a stable P1 helix is crucial for the binding of SAM to aptamer[151, 173], while our measurements indicate that stacking on an AT helix can to some degree substitute for a truncated segment of the P1 helix. We wished to determine
how SAM binding would be affected by the relative lengths of P1 and AT helices within the competition region shown in Figure 3.2. The P2_AT construct, which contains a full length AT helix but no P1 helix or J1/2, showed minimal binding with a B/A ratio of ~1.

**Figure 3.21**: 10% denaturing acrylamide gel showing the cleavage pattern in inline probing experiment for 3P1_9AT (A) and 3P1_10AT (B), 3P1_10AT_5’GCC. The lanes in the two gels from 1-5 represent No reaction, Partial alkaline digested RNA, RNase T1 digested RNA, cleaved RNA in inline probing buffer without SAM, cleaved RNA in inline probing buffer with 100 μM SAM. The arrows in red and blue indicate regions with enhanced and reduced cleavage.

The binding of SAM to the different hybrids represented by B/A ratio is within 2.5-4.5 in the equilibrium dialysis assay (Figure 3.22a) at RNA concentration of 3μM and SAM concentration of 100nM. Even 2P1_11AT, with a P1 helix as short as two base pairs, shows readily detectable binding to SAM. The 2P1_11AT construct nonetheless binds to SAM with weakest affinity amongst all the hybrid constructs. As the number of base pairs in the P1 helix is increased there is a small increase in SAM binding. The differences in $K_d$ are very close to standard deviation for multiple single point measurements, however.
Nonetheless the consistency of the trend of increasing affinity is consistent with a modest “long range” effect of P1 helix length on the stability of the SAM/riboswitch complex.

Figure 3.22: a) Ratio of number of counts for chamber b to chamber a in equilibrium dialysis for *B. subtilis* yitJ SAM-I riboswitch hybrid constructs (constructs constrained to form hybrids without any competing nucleotides) and the P2_AT construct in which the J1/2 region and the 5’ of the P1 helix are truncated. b) Ratio of number of counts for chamber B to chamber A in equilibrium dialysis for *B. subtilis* yitJ SAM-I riboswitch hybrid constructs with competing nucleotides (6P1_10AT, 8P1_11AT) and non base pairing nucleotides (3p_10at_5’GCC, 4P1_9AT5’GAC). C) Secondary structure of 4P1_9AT_5’GAC and 3P1_10AT_5’GCC constructs.

3.3.4.6. SAM Binding to the Hybrids Is Very Sensitive to Structural Defects Near the Binding Site

The predictions in Figure 3.16 suggest that under some conditions the riboswitch may form conformations that include the formation of hybrids containing partial P1 and
partial AT helices. We were therefore led to speculate as to whether the hybrid constructs shown in Figure 3.4b could be appropriate model systems for SAM binding to such intermediate conformations. Presumably the context of a riboswitch containing those 5’ residues that have been truncated in our constructs would present a different binding site geometry. Addition of non base pairing nucleotides on the 5’ end of the P1 helix represented by the 3P1_10AT_5’GCC construct decreased the SAM binding dramatically to a B/A ratio of 1.4 (Figure 3.22b) as opposed to a B/A ratio of 4.4 with 3P1_10AT. The inhibitory effect of adding non-pairing 5’ nucleotides is much less, however, for the 4P1_9AT_5’GAC construct Figure 3.22b). This value is half the value of SAM binding with 4P1_9AT (Figure 3.22a).

Competing nucleotides included from the 3’ end of the AT helix represented by 6P1_10_AT and 8P1_11AT constructs do not hinder SAM binding with a B/A ratio of ~4.01 and 4.3 respectively (Figure 3.22b). Thus, addition of 5’ nucleotides inhibits SAM binding if the 5’ overhang has two characteristics: 1) non-complementarity to residues on the opposing strand, and 2) the non-complementary residues are near the SAM binding site.

3.3.4.7. A Three-dimensional Model for SAM Binding to Hybrid ON/OFF SAM-I Riboswitch Conformation

Given that our equilibrium dialysis measurements suggested continuous stacking between P1 and AT helix segments, we constructed a three-dimensional model of 3P1_10AT.

As shown in Figure 3.24a, MC-sym can exhaustively sample the possible integration of the AT helix into the X-ray derived coordinates for the aptamer core without introducing
steric clashes. Then we used vdW interaction energy to monitor the co-axial stacking between the partial P1 helix and the incomplete AT helix. The stacking for two sets of nucleobases—U107-A108 and A1-U132 are monitored. The 2D histogram of the vdW energies is displayed in Figure 3.24b. The results show that MC-sym predicts some models with the co-axial stacking feature as demonstrated by the area with favorable vdW energies in both sets of nucleobases. One structural model with the co-axial helical stacking is shown in Figure 3.24c. The hybrid model (Figure 3.4) indicates that the partial AT helix can form a continuous stack with the partial P1 helix with little indication of steric clash with other segments of the aptamer structure, as predicted based upon the X-ray coordinates.

3.3.5. Discussion

3.3.5.1. The Basis for Discrimination Between Riboswitch Conformations by SAM Is More Complex than Suggested by Secondary Structure Diagrams

SAM riboswitches are an instructive example of nature using secondary and tertiary structure folding for specific molecular/metabolite recognition. Our work is aimed at understanding how RNA folding is integrated with metabolite binding to make a genetic decision. Functionally, the aptamer is considered to be the biosensing domain of the riboswitch, utilizing 3D structural motifs [108, 168, 174-177]. Five classes of SAM riboswitch aptamers recognize SAM specifically-each using distinct binding pockets and structural motifs [178]. The expression platform undergoes large secondary structure changes to execute the gene expression controlling decision.
Figure 3.24: (a) Superposition of models generated from MC-sym sampling for hybrid construct 3P1-10AT. The region shown in blue is from the crystal structure of the yitJ SAM-I riboswitch (PDB:3NPB) [168]. The part in red is the AT helix from MC-sym modeling. (b) 2D histogram of the vDW interaction energies between the nucleobase of U107 and that of A108 and between the nucleobase of A1 and that of U132. A more negative value of vDW energy here indicates better stacking between the nucleobases. (c) An example of models sampling the co-axial helical stacking between the partial P1 and the AT helix with SAM docked into the binding pocket based on the binding mode observed in the aptamer crystal structure. SAM is shown in vDW representation with carbon colored in yellow. The partial P1, J4/1 and P4 helix are shown in blue, J1/2 in orange, J3/4 in magenta and P2 & P3 in green. The modeled AT helix is shown in red.

In the case of the SAM-I riboswitch the aptamer domain has been well characterized [144, 175, 178]. It has been more challenging to devise strategies to study the expression platform because of its high sequence variability [143]. Secondary structure models for most riboswitch On states differ dramatically from two-dimensional representations of the Off state [106, 151, 178, 179]. For the SAM-I (and many other) riboswitches however, the switching mechanism requires that the base compositions of the P1 and AT helices be
similar. In that case, one cannot dismiss the possibility of an On state secondary structure that mimics some aspects of the Off state in its tertiary folding, with the AT helix being positioned in a manner to replace the P1 helix. Moreover, RNA structure, like protein structure, is more accurately described in solution as a Boltzmann distribution of conformer populations, rather than a single structure [180, 181]. This may be especially true of unliganded riboswitches, which are the product of evolutionary selection for conformational dynamics. Therefore riboswitches are likely to sample conformers with “hybrid” On and Off state character, with the interchange between them facilitated if tertiary folds are similar.

These considerations led us to test the SAM binding properties of a set of SAM-I riboswitch constructs constrained to form likely conformational intermediates. Truncating and mutating riboswitches and other folded RNAs has been a very effective means of constraining conformational dynamics to facilitate biophysical studies [178, 182, 183]. This strategy enabled us to detect a reduced level of binding to conformational substrates that will be transiently present in partly or fully transcribed SAM-I riboswitch domains in solution.

3.3.5.2. SAM Recognizes Similar Tertiary Characteristics in Hybrid and Aptamer SAM-I Riboswitch RNA Segments

We observed SAM binding to all RNA constructs that contained a partial P1 helix, as long as a potential AT helix extending to the P1 helix boundary was present (Figures 3.19-3.22). The kink-turn mediated pseudoknot interaction acts as an auxiliary for the global folding in the native SAM aptamer structure leading to a tight compaction upon SAM
binding [152, 172]. According to our measurements for 3P1_AT_P2 mutant and its compensatory construct (Figure 3.22), similar dependence on the pseudoknot for SAM binding is seen in the hybrids. The magnesium effect on the aptamer is previously observed from FRET studies, gel shift assays, and chemical probing [151, 152, 172]. The magnesium dependence observed in the hybrids (Figure 3.19) is similar to the magnesium dependence observed in the isolated aptamer construct. The magnesium based SAM binding data and the pseudoknot mutational analysis in the hybrids indicate that the junction regions and the helices in the aptamer region of the hybrids retain the global folding and SAM binding mode observed in the isolated aptamer constructs.

We have constructed a model of 3P1_AT, which shows that an AT helix can be accommodated within a global architecture resembling the aptamer fold observed in X-ray structures (Figure 3.24). The model has been constructed using the X-ray coordinates for all elements outside of the AT helix. MD simulations indicate that this configuration with SAM bound is relatively stable, but that the P1 helix base pairs dissociate in the absence of SAM (Huang et al, manuscript in preparation).

3.3.5.3. Minimal P1 Helices Required for SAM Recognition can be Stabilized in Hybrids by Coaxially-stacked AT Helices

Our hybrid constructs of the *B.subtilis yitJ* SAM-I riboswitch include the AT helix which is part of the expression platform. The SAM binding data with the hybrids presented here suggests that aptamer is not the only domain that can affect SAM binding and recognition. Our data suggests that the nucleotides in the AT helix can, up to a point, substitute for P1 helix base pairs in SAM recognition. Recent studies with the SAM-I riboswitch showed
that aptamers having 3 or 4 or even 6 base pairs in the P1 helix in aptamer constructs do not terminate transcription as a result of addition of 10μM SAM [172]. This is in agreement with our SAM binding data with the 3P1 only construct (Figure 3.19), since transcription termination is associated with SAM-induced P1 helix formation. The hybrid constructs 2P1_11AT, 3P1_10AT, 4P1_9AT, 5P1_8AT and 6P1_7AT all bind to SAM, though with reduced affinity compared to the aptamer (Figures 3.18-3.22). Thus the AT can contribute to SAM recognition along with the remaining crucial components of the aptamer. In particular, in addition to helices P2-P4 and the pseudoknot interaction, a P1 helix nucleation site and constrained J1/2 are required, since binding is not detectable for P2_AT (Figure 3.19). It is worth noting that the first two base pairs of the P1 helix and residues in J1/2 are involved in critical contacts with SAM according to X-ray data [107, 108, 168, 178] and MD simulations [184].

The SAM binding results with the 3P1 only, 3P1_9AT (Figure 3.19) suggest that the 3P1 is stabilized by the AT helix through putative stacking interaction between the two helices (also seen from the MD simulations). Thus, in the absence of a continuously stacked AT helix, the shortened P1 helices do not form and no SAM binding pocket is present. Our data with the 3P1_10AT5’GCC (Figure 3.22) show that the 5’GCC3’ sequence on the 5’ end of the P1 helix in the 3P1_10AT hybrid construct inhibits SAM binding. This destabilization may be due to disruption of the P1/AT coaxial stack, leading to a dissociation of the three P1 helix base pairs. Alternatively, it may arise from steric interference between the 5’ overhanging nucleotides and the ligand or its binding pocket.

When 1 base pairing nucleotide is added on the P1 helix represented by the 4P1_9AT5’GAC, SAM binding is inhibited relative to 4P1_9AT (Figure 3.22), but far less
inhibition is observed than when a similar substitution of unpaired 5’ nucleotides is added to the 3P1_10AT to make 3P1_10AT5’GCC. As the length of the P1 helix increases, the stability of the P1 helix increases, but steric clash with the binding site may also be reduced.

These results also hint at a dynamic mechanism in which, upon the formation of each base pair in the P1 helix, the equilibrium is forced towards the Off state. We can hypothesize that this mechanism may be present in the constructs 6P1_10AT and 8P1_11AT, which contain the potential for competing P1 and AT helices. While we cannot say which, if any, conformers predominate in these two constructs in the absence of ligand, we suggest that SAM binding will push the equilibrium towards those states that contain longer P1 helices. This mechanism is bolstered by the increasing SAM binding trend, however weak, observed with the increasing length of the P1 helix in the hybrids. The apparent interference of 5’ overhangs near the SAM binding pocket with SAM binding will push the equilibrium even more strongly in the direction of an extended P1. Our MD simulations with 6P1_10AT indicate a “branch migration” event in the presence of SAM (Huang et al, manuscript in preparation). Further research needs to be done in the AT region in conjunction with the P1 helix to pin down the mechanistic aspects of the P1 and AT stacking and branch migration.

3.3.5.4. The Role of Conformational Dynamics and Conformational Selection in SAM Binding

Earlier studies indicated that an ensemble of conformational states exists in the unliganded SAM-I riboswitch aptamer [151, 173]. SAM selectively prefers some of these
intermediate conformations for binding. The extension of the riboswitch transcript to include 3’ nucleotides with the potential to compete with P1 helix base pairing should increase the conformational heterogeneity of the conformer population. In solution, 2P1_11AT, 3P1_10AT, 4P1_9AT, 5P1_8AT, 6P1_7AT and 8P1_5AT represent SAM-selected conformational species, along with the aptamer. The hybrids 3P1 ONLY, P2-AT, 3P1_9AT represent conformational species of the SAM-I riboswitch that are not preferred by SAM for binding. With stronger affinity for the aptamer (forming the full P1 helix) than for the hybrids, SAM would rapidly shift the equilibrium towards the Off state. We do not have an explanation as to why the constructs 8P1_5AT and 8P1_11AT, which contain the potential to form full P1 helices, bind SAM with weaker affinity than the truncated aptamer (Figure 6). Nonetheless, the implication of the latter finding is that SAM binding can eventually destabilize full AT helix formation, thus facilitating the downstream formation of a rho-independent terminator hairpin. Our rational design of the hybrids reveals the elements of the SAM-I riboswitch that confer the ability to recognize and bind SAM beyond the aptamer domain.

3.3.5.5. Pathways for the SAM-induced Riboswitch Conformational Folding During Transcription

It has been shown that some riboswitches operate on a kinetic and/or thermodynamic control mechanism [185, 186]. The SAM-I riboswitch is a transcriptional riboswitch. The timescale at which the decision is to be made between the formation of either conformer is in seconds considering the processivity of prokaryotic RNAP [67]. Whether this timescale is sufficient for the growing transcript to fold under conditions of
thermodynamic equilibrium depends on rates for folding and ligand binding. It will also be affected by the presence of proteins and other components of the transcription complex. It has been suggested that, in the presence of SAM, the decision to form the aptamer conformation, and therefore to terminate transcription, is fixed by SAM binding before the AT helix is transcribed [178, 179]. Our findings are consistent with this hypothesis. If SAM is bound to the partial riboswitch transcript containing the aptamer conformation with full P1 helix, the 20-100 fold preference of the ligand for the aptamer over AT-helix containing conformers will prevent AT helix formation.

At the same time, our results also suggest the possibility of a second path for SAM-induced conformational switching. Folding of the On conformer, which consists of proximally-formed helical base pairs, may be more rapid than the formation of distal base pairs required for the P1 helix [184, 187]. This may be especially true if the latter is slowed by transient pairing of purine residues from the junction regions with 5’ pyrimidine residues from the P1 helix (Huang et al, manuscript in preparation). Moreover, literature reports suggest that before SAM binding P1 helix formation will be somewhat dynamic [172, 173]. If AT helix formation can stabilize the nucleation of a nascent P1 helix, then our results show that SAM could bind to the resulting hybrid conformation. Furthermore, SAM binding should result in rapid conversion to a fully formed P1 helix and freeing of potential AT-forming residues to participate in terminator formation. One attractive aspect of this model is that it predicts that termination will require μM concentrations of SAM to stabilize the transition state. In vitro transcription assays with the *yijf* SAM-I riboswitch do indicate that much higher levels of SAM are required for
transcription termination than would be predicted based upon in vitro equilibrium measurements of SAM binding affinity [179].

Distinguishing between these two pathways requires consideration of the complex interplay between SAM binding, RNA folding, transcriptional pausing, and the role of the transcription complex itself.

3.4. Conclusions

The SAM binding studies with the “alternative” structures mentioned in this paper indicate three major aspects of the SAM-I riboswitch folding responsible for ligand discrimination between the two functional conformers of the riboswitch 1) Stable P1 helix is necessary for the SAM binding and recognition, 2) the AT helix can play a role in the transition from the On state to the Off state of the riboswitch by stabilizing the P1 helix with stacking interactions. As the length of the P1 helix increases, the stability of the P1 helix increases, and the affinity for SAM increases. 3) That SAM-binding affinity for the hybrid conformers is weaker by 1-2 orders of magnitude compared to that for the aptamer implies that SAM binding to hybrids would rapidly convert them to the Off state.

Our study indicates that SAM can bind to a SAM-I riboswitch segment which is constrained to form a full length antiterminator (2P1_11AT), and also to hybrid constructs which, within the context of a full-length riboswitch, would block terminator formation. This result may appear counterintuitive. Nonetheless, the mechanism for SAM-I riboswitch function is readily compatible with our findings. On the other hand, our results show the minimal secondary structure requirements to form a SAM-binding competent tertiary fold, for SAM-I riboswitches. Other riboswitches, such as the purine
[188] and lysine [90] riboswitches, as well as simpler systems such as the quenosine [189] and SAMIII-SMK [190] riboswitch, also contain helix-competing motifs. It remains to be seen what role may be played by conformational intermediates, which combine Off state tertiary structure characteristics with On state secondary structure elements in the function of SAM-I and other riboswitches.
CHAPTER 4: SUMMARY AND FUTURE DIRECTIONS

This work focuses on understanding the role of a small molecule in RNA folding and the basis for RNA-small molecule recognition (T.tencongensis metF SAM-I riboswitch and the B.subtilis yitJ SAM-I riboswitch and the RNA directed small molecule library). The SAM-I riboswitch is known to be involved in controlling the genes responsible for virulence in certain pathogenic bacteria (See chapter 1). The two SAM-I riboswitches mentioned in this work recognize a ligand (SAM) with high specificity coupled with conformational changes in the RNA to modulate gene expression. Though competing P1 and AT helices were identified earlier, structural research on the SAM-I riboswitch done prior to this have not addressed the role of the AT helix in the gene regulation mechanism. Our work with the SAM-I riboswitch extends the structural studies into the AT helix and reveals the elements necessary for SAM to distinguish between the On and Off states of the B. Subtilis yitJ SAM-I riboswitch. Our work is motivated towards the hypothesis that SAM could bind weakly to the hybrids or On constructs derived from the identification of the base pairing pattern similarity between the AT and P1 helices in both the riboswitches.

We used ligand binding detection techniques like NMR, equilibrium dialysis and inline probing. NMR is very sensitive to the detection of weak binding and an effective tool to probe the dynamics of a macromolecule and its interactions with small molecules and vice versa. In our case, the SAM-I riboswitch being a large molecule for NMR, with overlapping peaks, we had difficulties in optimizing RNA-detected SAM binding studies useful for structure analysis. However we have been successful in detecting SAM binding.
using NMR with ligand detected experiments. We used NMR to show that the SAM-I riboswitch aptamer and an On construct induces changes in ligand peaks.

Equilibrium dialysis is highly sensitive to ligand binding and can be effectively used for quantification of the binding affinity. We used equilibrium dialysis to detect and quantify SAM binding for the SAM-I riboswitch constructs. In the hybrids, we could see a minor increase in the binding to SAM as the length of the P1 helix is increased. Previously, the apparent $K_d$ with SAM was determined to be within a range of 4nM-30nM for the aptamer only (with 8 base pairs in the P1 helix, without antiterminator residues) and the full length SAM-I riboswitch with the residues involved in the formation of the terminator as well, in which the AT is not expected to form in the SAM bound configuration. In our study with the hybrids, we observed an apparent $K_d$ of $\sim500nM$ (8P1_5AT). We could not see “full” binding in the 8P1_5AT as seen in the aptamer or in the full-length riboswitch.

Geometric/steric interference caused by the AT helix could be restricting the P1 helix from binding to SAM optimally. Another explanation for the decreased SAM binding in the 8P1_5AT could be that the bulge in the AT helix causes slippage in AT helix base pairing and causing instability at the P1/AT interface in the 8P1_5AT construct (Figure 4.1). This leads to the hypothesis that the AT is necessary to stabilize the partially formed P1 helix (as seen from the data presented) yet could be involved in controlling the terminator formation based on the physiological SAM concentration. One experiment we propose is to test with 8P1_4AT. If the SAM binding increases, by decreasing the number of AT base pairs, the “strand switching” mechanism in which the AT helix “unzips” allowing SAM to completely bind to the aptamer and make the residues of the AT available for terminator formation can be plausible for this riboswitch. Ultimately a crystal structure of the
8P1_5AT will give the insights into the role of the AT helix and the single base (U) bulge. In order to crystallize the 8P1_5AT, the AT helix loop has to be modified into a tetraloop. Hence another 8P1_5AT with a tetraloop has to be tested for SAM binding to obtain similar SAM binding level as with the original 8P1_5AT hybrid.

**Figure 4.1:** A secondary structure representation for 8P1_5AT and the probable slippage caused by the u bulge in the AT shown at the opposite end of the arrow.

The SAM binding data from our mutations in the J1/2, P2 helix in the hybrids show that the global folding of the riboswitch by the Kink-turn-assisted Pseudoknot with tertiary interactions is similar to that of the aptamer. The A109 and U86 in the J4/1 and the J3/4, respectively, form a base pair in the aptamer as seen from the crystal structure [108]. Mutation analysis of these two residues will further give insights into global folding of the
hybrids or alternatively a high-resolution crystal structure of any hybrid will give more insights into the global folding of the riboswitch in the hybrid configuration and also reveal the interhelical stacking interactions between the P1 helix and the AT helix. The crystal structure will show the effect of a stable complete P1 helix and P1/AT hybrid helices in the long-range interactions mentioned above along with more insights into the folding of the RNA to attain high affinity for SAM binding.

The inline probing shows the distinction between the SAM binding observed in the 3P1_10AT and the 3p1_10AT_5'GCC and 3P1_9AT constructs.

Figure 4.1: A model for SAM-I riboswitch folding in the presence of SAM. The blue line and the dashed blue line indicate tertiary interactions. The shaded region indicates putative P1 helix nucleation region.
Our results that the SAM-I riboswitch hybrids bind to SAM open a new venue for riboswitch drug targeting. More structural information on the SAM bound hybrids will be useful in developing high affinity ligands to the hybrids. The ligand bound hybrid can stabilize the riboswitch in the hybrid configuration and not allow the formation of the P1 helix and remain in a permanent On state or in the hybrid state with weak expression of the genes. Rationally it will be more difficult to design ligands to the aptamer itself to shutdown the gene expression since the aptamer has evolved with high specificity for SAM. The hybrids could form “alternate” binding pocket that can accommodate a ligand other than SAM. All the aforementioned experiments can be tested in vitro [88], and functional screens can be used to test the role of hybrids in vivo if ligand binding to the hybrids are identified.

The RNA directed library is constructed specifically to assemble ligands with RNA bindability and understand the modes of novel RNA-ligand binding properties from the obtained hits. Though we obtained 2 new hits from our library against, the 16s ribosomal RNA that is previously shown to bind to some aminoglycoside antibiotics, more detailed structure of the bound ligand is required for understanding the desired properties of RNA. Also the larger RNA targets like the riboswitch have to be screened. Large RNAs are folded through long-range tertiary interactions for function and can potentially form binding pockets for ligands. We tried to optimize the screening parameters for screening the yitJ SAM-I riboswitch using NMR and equilibrium dialysis. Since it is shown that the S-adenosyl homocysteine binds to the SAM-I riboswitch with 400nM apparent Kₐ and S-adenosyl cysteine with ~30 micromolar apparent Kₐ that fall in the weak binding
category (as expected for the fragment’s affinity from the library screen) these ligands can be used for optimizing the NMR directed binding experiments.

In summary, we have used various methods (equilibrium dialysis, RNA detected NMR, ligand detected NMR, inline probing, mutational analysis) to understand the role of a small molecule in RNA binding and folding. Our findings can be applicable and useful for further RNA targeted drug development against pathogenic organisms.
REFERENCES


APPENDIX: COPYRIGHT PERMISSIONS FROM JOURNALS

NATURE PUBLISHING GROUP LICENSE
TERMS AND CONDITIONS

Nov 03, 2011

This is a License Agreement between vamsi K boyapati ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

License Number 2781600826386
License date Nov 03, 2011
Licensed content publisher Nature Publishing Group
Licensed content publication Nature Biotechnology
Licensed content title Riboswitches as antibacterial drug targets
Licensed content author Kenneth F Blount and Ronald R Breaker
Licensed content date Dec 1, 2006
Volume number 24
Issue number 12
Type of Use reuse in a thesis/dissertation
Requestor type academic/educational
Format print and electronic
Portion figures/tables/illustrations
Number of figures/tables/illustrations 1
High-res required no
Figures Figure 2
Author of this NPG article no
Your reference number
Title of your thesis / dissertation STUDY OF COMPLEX RNA FUNCTION MODULATED BY SMALL MOLECULES: THE DEVELOPMENT OF RNA DIRECTED SMALL MOLECULE LIBRARY AND PROBING THE S-ADENOSYL METHIONINE DISCRIMINATION BETWEEN ON ANDOFF CONFORMATIONAL STATES OF THE SAM-I RIBOSWITCH.
Expected completion date Jan 2012
Estimated size (number of pages) 100
Total 0.00 USD
Terms and Conditions

Rightslink Printable License https://s100.copyright.com/App/PrintableLicenseFrame.jsp?pub...
Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version. Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run). NB: In all cases, web-based use of full-text articles must be authorized separately through the ‘Use on a Web Site’ option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

   **Note:** For re-publication from the British Journal of Cancer, the following credit lines apply:
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)
   For AOP papers, the credit line should read:
   Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:
   Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   **Note:** For adaptation from the British Journal of Cancer, the following credit line applies:
   Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information. Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:
   Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

   **Note:** For translation from the British Journal of Cancer, the following credit line applies:
   Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.

Special Terms:

v1.1
This is a License Agreement between vamsi K boyapati ("You") and Nature Publishing Group ("Nature Publishing Group") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Nature Publishing Group, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>License Number</th>
<th>2781610288861</th>
</tr>
</thead>
<tbody>
<tr>
<td>License date</td>
<td>Nov 03, 2011</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Nature Publishing Group</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Nature Chemical Biology</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Antibacterial lysine analogs that target lysine riboswitches</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Kenneth F Blount, Joy Xin Wang, Jinsoo Lim, Narasimhan Sudarsan and Ronald R Breaker</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>Dec 3, 2006</td>
</tr>
<tr>
<td>Volume number</td>
<td>3</td>
</tr>
<tr>
<td>Issue number</td>
<td>1</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a thesis/dissertation</td>
</tr>
<tr>
<td>Requestor type</td>
<td>academic/educational</td>
</tr>
<tr>
<td>Format</td>
<td>print and electronic</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables/illustrations</td>
<td>1</td>
</tr>
<tr>
<td>High-res required</td>
<td>no</td>
</tr>
<tr>
<td>Figures</td>
<td>figure 2</td>
</tr>
<tr>
<td>Author of this NPG article</td>
<td>no</td>
</tr>
<tr>
<td>Your reference number</td>
<td></td>
</tr>
<tr>
<td>Title of your thesis/dissertation</td>
<td>STUDY OF COMPLEX RNA FUNCTION MODULATED BY SMALL MOLECULES: THE DEVELOPMENT OF RNA DIRECTED SMALL MOLECULE LIBRARY AND PROBING THE S-ADENOSYL METHIONINE DISCRIMINATION BETWEEN ON AND OFF CONFORMATIONAL STATES OF THE SAM-I RIBOSWITCH.</td>
</tr>
<tr>
<td>Expected completion date</td>
<td>Jan 2012</td>
</tr>
<tr>
<td>Estimated size (number of pages)</td>
<td>100</td>
</tr>
<tr>
<td>Total</td>
<td>0.00 USD</td>
</tr>
</tbody>
</table>
Terms and Conditions

Terms and Conditions for Permissions

Nature Publishing Group hereby grants you a non-exclusive license to reproduce this material for this purpose, and for no other use, subject to the conditions below:

1. NPG warrants that it has, to the best of its knowledge, the rights to license reuse of this material. However, you should ensure that the material you are requesting is original to Nature Publishing Group and does not carry the copyright of another entity (as credited in the published version). If the credit line on any part of the material you have requested indicates that it was reprinted or adapted by NPG with permission from another source, then you should also seek permission from that source to reuse the material.

2. Permission granted free of charge for material in print is also usually granted for any electronic version of that work, provided that the material is incidental to the work as a whole and that the electronic version is essentially equivalent to, or substitutes for, the print version. Where print permission has been granted for a fee, separate permission must be obtained for any additional, electronic re-use (unless, as in the case of a full paper, this has already been accounted for during your initial request in the calculation of a print run). NB: In all cases, web-based use of full-text articles must be authorized separately through the 'Use on a Web Site' option when requesting permission.

3. Permission granted for a first edition does not apply to second and subsequent editions and for editions in other languages (except for signatories to the STM Permissions Guidelines, or where the first edition permission was granted for free).

4. Nature Publishing Group's permission must be acknowledged next to the figure, table or abstract in print. In electronic form, this acknowledgement must be visible at the same time as the figure/table/abstract, and must be hyperlinked to the journal's homepage.

5. The credit line should read:
Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)
For AOP papers, the credit line should read:
Reprinted by permission from Macmillan Publishers Ltd: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

Note: For republication from the British Journal of Cancer, the following credit lines apply.
Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication) For AOP papers, the credit line should read:
Reprinted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME], advance online publication, day month year (doi: 10.1038/sj.[JOURNAL ACRONYM].XXXXX)

6. Adaptations of single figures do not require NPG approval. However, the adaptation should be credited as follows:
Adapted by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

Note: For adaptation from the British Journal of Cancer, the following credit line applies:
Adapted by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

7. Translations of 401 words up to a whole article require NPG approval. Please visit http://www.macmillanmedicalcommunications.com for more information. Translations of up to a 400 words do not require NPG approval. The translation should be credited as follows:
Translated by permission from Macmillan Publishers Ltd: [JOURNAL NAME] (reference citation), copyright (year of publication)

Note: For translation from the British Journal of Cancer, the following credit line applies:
Translated by permission from Macmillan Publishers Ltd on behalf of Cancer Research UK: [JOURNAL NAME] (reference citation), copyright (year of publication)

We are certain that all parties will benefit from this agreement and wish you the best in the use of this material. Thank you.
## ELSEVIER LICENSE
### TERMS AND CONDITIONS

Nov 09, 2011

This is a License Agreement between vamsi K boyapati ("You") and Elsevier ("Elsevier") provided by Copyright Clearance Center ("CCC"). The license consists of your order details, the terms and conditions provided by Elsevier, and the payment terms and conditions.

All payments must be made in full to CCC. For payment instructions, please see information listed at the bottom of this form.

<table>
<thead>
<tr>
<th>Supplier</th>
<th>Elsevier Limited</th>
</tr>
</thead>
<tbody>
<tr>
<td>Registered Company Number</td>
<td>1982084</td>
</tr>
<tr>
<td>Customer name</td>
<td>vamsi K boyapati</td>
</tr>
<tr>
<td>Customer address</td>
<td>401 choppin hall</td>
</tr>
<tr>
<td></td>
<td>baton rouge, LA 70808</td>
</tr>
<tr>
<td>License number</td>
<td>2781611354115</td>
</tr>
<tr>
<td>License date</td>
<td>Nov 03, 2011</td>
</tr>
<tr>
<td>Licensed content publisher</td>
<td>Elsevier</td>
</tr>
<tr>
<td>Licensed content publication</td>
<td>Bioorganic &amp; Medicinal Chemistry Letters</td>
</tr>
<tr>
<td>Licensed content title</td>
<td>Rational design of SAM analogues targeting SAM-II riboswitch aptamer</td>
</tr>
<tr>
<td>Licensed content author</td>
<td>Young Wan Ham, Darin J. Humphreys, Susie Choi, Daniel L. Dayton</td>
</tr>
<tr>
<td>Licensed content date</td>
<td>1 September 2011</td>
</tr>
<tr>
<td>Licensed content volume number</td>
<td>21</td>
</tr>
<tr>
<td>Licensed content issue number</td>
<td>17</td>
</tr>
<tr>
<td>Number of pages</td>
<td>4</td>
</tr>
<tr>
<td>Start Page</td>
<td>5071</td>
</tr>
<tr>
<td>End Page</td>
<td>5074</td>
</tr>
<tr>
<td>Type of Use</td>
<td>reuse in a thesis/dissertation</td>
</tr>
<tr>
<td>Intended publisher of new work</td>
<td>other</td>
</tr>
<tr>
<td>Portion</td>
<td>figures/tables/illustrations</td>
</tr>
<tr>
<td>Number of figures/tables /illustrations</td>
<td>1</td>
</tr>
</tbody>
</table>
STUDY OF COMPLEX RNA FUNCTION MODULATED BY SMALL MOLECULES: THE DEVELOPMENT OF RNA DIRECTED SMALL MOLECULE LIBRARY AND PROBING THE S-ADENOSYL METHIONINE DISCRIMINATION BETWEEN ON AND OFF CONFORMATIONAL STATES OF THE SAM-I RIBOSWITCH.
8. License Contingent Upon Payment: While you may exercise the rights licensed immediately upon issuance of the license at the end of the licensing process for the transaction, provided that you have disclosed complete and accurate details of your proposed use, no license is finally effective unless and until full payment is received from you (either by publisher or by CCC) as provided in CCC's Billing and Payment terms and conditions. If full payment is not received on a timely basis, then any license preliminarily granted shall be deemed automatically revoked and shall be void as if never granted. Further, in the event that you breach any of these terms and conditions or any of CCC's Billing and Payment terms and conditions, the license is automatically revoked and shall be void as if never granted. Use of materials as described in a revoked license, as well as any use of the materials beyond the scope of an unretracted license, may constitute copyright infringement and publisher reserves the right to take any and all action to protect its copyright in the materials.

9. Warranties: Publisher makes no representations or warranties with respect to the licensed material.

10. Indemnity: You hereby indemnify and agree to hold harmless publisher and CCC, and their respective officers, directors, employees and agents, from and against any and all claims arising out of your use of the licensed material other than as specifically authorized pursuant to this license.

11. No Transfer of License: This license is personal to you and may not be sublicensed, assigned, or transferred by you to any other person without publisher's written permission.

12. No Amendment Except in Writing: This license may not be amended except in a writing signed by both parties (or, in the case of publisher, by CCC on publisher's behalf).

13. Objection to Contrary Terms: Publisher hereby objects to any terms contained in any purchase order, acknowledgment, check endorsement or other writing prepared by you, which terms are inconsistent with these terms and conditions or CCC's Billing and Payment terms and conditions. These terms and conditions, together with CCC's Billing and Payment terms and conditions (which are incorporated herein), comprise the entire agreement between you and publisher (and CCC) concerning this licensing transaction. In the event of any conflict between your obligations established by these terms and conditions and those established by CCC's Billing and Payment terms and conditions, these terms and conditions shall control.

14. Revocation: Elsevier or Copyright Clearance Center may deny the permissions described in this License at their sole discretion, for any reason or no reason, with a full refund payable to you. Notice of such denial will be made using the contact information provided by you. Failure to receive such notice will not alter or invalidate the denial. In no event will Elsevier or Copyright Clearance Center be responsible or liable for any costs, expenses or damage incurred by you as a result of a denial of your permission request, other than a refund of the amount(s) paid by you to Elsevier and/or Copyright Clearance Center for denied permissions.

**LIMITED LICENSE**

The following terms and conditions apply only to specific license types:

15. **Translation**: This permission is granted for non-exclusive world English rights only unless your license was granted for translation rights. If you licensed translation rights you may only translate this content into the languages you requested. A professional translator must perform all translations and reproduce the content word for word preserving the integrity of the article. If this license is to re-use 1 or 2 figures then permission is granted for non-exclusive world rights in all languages.

16. **Website**: The following terms and conditions apply to electronic reserve and author websites:

   **Electronic reserve**: If licensed material is to be posted to website, the web site is to be password-protected and made available only to bona fide students registered on a relevant course if:
   
   This license was made in connection with a course.
   
   This permission is granted for 1 year only. You may obtain a license for future website posting.
   
   All content posted to the web site must maintain the copyright information line on the bottom of each image, A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx or the Elsevier homepage for books at http://www.elsevier.com , and Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

17. **Author website** for journals with the following additional clauses:

   All content posted to the web site must maintain the copyright information line on the bottom of each image, and the permission granted is limited to the personal version of your paper. You are not allowed to download and post the published electronic version of your article (whether PDF or HTML, proof or final version), nor may you scan the printed edition to create an electronic version, A hyper-text must be included to the Homepage of the journal from which you are licensing at
As part of our normal production process, you will receive an e-mail notice when your article appears on Elsevier’s online service ScienceDirect (www.sciencedirect.com). That e-mail will include the article’s Digital Object Identifier (DOI). This number provides the electronic link to the published article and should be included in the posting of your personal version. We ask that you wait until you receive this e-mail and have the DOI to do any posting.

Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

18. **Author website** for books with the following additional clauses:
   - Authors are permitted to place a brief summary of their work online only.
   - A hyper-text must be included to the Elsevier homepage at http://www.elsevier.com
   - All content posted to the web site must maintain the copyright information line on the bottom of each image
   - You are not allowed to download and post the published electronic version of your chapter, nor may you scan the printed edition to create an electronic version.
   - Central Storage: This license does not include permission for a scanned version of the material to be stored in a central repository such as that provided by Heron/XanEdu.

19. **Website** (regular and for author): A hyper-text must be included to the Homepage of the journal from which you are licensing at http://www.sciencedirect.com/science/journal/xxxxx, or for books to the Elsevier homepage at http://www.elsevier.com

20. **Thesis/Dissertation**: If your license is for use in a thesis/dissertation your thesis may be submitted to your institution in either print or electronic form. Should your thesis be published commercially, please reapply for permission. These requirements include permission for the Library and Archives of Canada to supply single copies, on demand, of the complete thesis and include permission for UMI to supply single copies, on demand, of the complete thesis. Should your thesis be published commercially, please reapply for permission.

21. **Other Conditions**:

v1.6

If you would like to pay for this license now, please remit this license along with your payment made payable to "COPYRIGHT CLEARANCE CENTER" otherwise you will be invoiced within 48 hours of the license date. Payment should be in the form of a check or money order referencing your account number and this invoice number RLNS00658207.

Once you receive your invoice for this order, you may pay your invoice by credit card. Please follow instructions provided at that time.

Make Payment To:
Copyright Clearance Center
Dept 001
P.O. Box 843006
Boston, MA 02284-3006

For suggestions or comments regarding this order, contact RightsLink Customer Support: customercare@copyright.com or +1-877-622-5543 (toll free in the US) or +1-978-646-2777.

Gratis licenses (referencing $0 in the Total field) are free. Please retain this printable license for your reference. No payment is required.
Title: Design and Antimicrobial Action of Purine Analogue That Bind Guanine Riboswitches
Author: Jane N. Kim et al.
Publication: ACS Chemical Biology
Publisher: American Chemical Society
Date: Nov 1, 2009

PERMISSION/LICENSE IS GRANTED FOR YOUR ORDER AT NO CHARGE

This type of permission/license, instead of the standard Terms & Conditions, is sent to you because no fee is being charged for your order. Please note the following:

- Permission is granted for your request in both print and electronic formats.
- If figures and/or tables were requested, they may be adapted or used in part.
- Please print this page for your records and send a copy of it to your publisher/graduate school.
- Appropriate credit for the requested material should be given as follows: "Reprinted (adapted) with permission from (COMPLETE REFERENCE CITATION). Copyright (YEAR) American Chemical Society." Insert appropriate information in place of the capitalized words.
- One-time permission is granted only for the use specified in your request. No additional uses are granted (such as derivative works or other editions). For any other uses, please submit a new request.
VITA

Vamsi Krishna Boyapati was born in August 1982, in Hyderabad, Andhra Pradesh, India. He received his Bachelor of Science in 2003 from Osmania University, Andhra Pradesh, India and his Master of Science degree from Acharya Nagarjuna University, Andhra Pradesh, India. He joined Dr. Fareed Aboul-Ela’s lab at Louisiana State University, Baton Rouge, Louisiana in January 2006. Vamsi Krishna Boyapati will graduate with the degree of Doctorate of Philosophy in December 2011.