A STRUCTURAL ANALYSIS OF A SAM-I/IV VARIANT RIBOSWITCH

Phillip E. Ross
Batey Laboratory

A thesis submitted to the Honors Council of the Department of Chemistry/Biochemistry of the University of Colorado in partial fulfillment of the requirements of the degree of Bachelor of Arts in Chemistry/Biochemistry with Honors.

Department of Chemistry/Biochemistry

Defense Date: April 1, 2013

Thesis Advisor: Dr. Robert Batey

Defense Committee:

Dr. Robert Batey (CHEM/BIOCHEM)
Dr. Robin Knight (CHEM/BIOCHEM)
Dr. Cynthia Carey (IPHY)
TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>TOPIC</th>
<th>PAGE NUMBER</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>3</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td>4</td>
</tr>
<tr>
<td>INTRODUCTION</td>
<td></td>
</tr>
<tr>
<td>Gene Regulation</td>
<td>5</td>
</tr>
<tr>
<td>Riboswitches</td>
<td>7</td>
</tr>
<tr>
<td>Structure</td>
<td>8</td>
</tr>
<tr>
<td>Applications</td>
<td>12</td>
</tr>
<tr>
<td>S-Adenosylmethionine Binding Riboswitches</td>
<td>15</td>
</tr>
<tr>
<td>SAM-I Riboswitch</td>
<td>16</td>
</tr>
<tr>
<td>Similarities between SAM-I and SAM-IV Riboswitches</td>
<td>18</td>
</tr>
<tr>
<td>SAM-I/IV Riboswitch</td>
<td>19</td>
</tr>
<tr>
<td>Research Objective</td>
<td>20</td>
</tr>
<tr>
<td>METHODS</td>
<td></td>
</tr>
<tr>
<td>DNA Design/Synthesis</td>
<td>23</td>
</tr>
<tr>
<td>RNA Synthesis</td>
<td>24</td>
</tr>
<tr>
<td>RNA Purification</td>
<td>25</td>
</tr>
<tr>
<td>Crystallization Condition Screen</td>
<td>26</td>
</tr>
<tr>
<td>X-Ray Diffraction</td>
<td>27</td>
</tr>
<tr>
<td>Molecular Modeling</td>
<td>28</td>
</tr>
<tr>
<td>Isothermal Titration Calorimetry</td>
<td>28</td>
</tr>
<tr>
<td>ITC Sample Preparation</td>
<td>30</td>
</tr>
<tr>
<td>RESULTS AND DISCUSSION</td>
<td></td>
</tr>
<tr>
<td>Structure Analysis of SAM-I/IV</td>
<td>31</td>
</tr>
<tr>
<td>Crystallography of SAM-I/IV Mutants</td>
<td>31</td>
</tr>
<tr>
<td>3’U Overhang</td>
<td>33</td>
</tr>
<tr>
<td>3’UC, 3’G Overhang, and no 3’U Overhang</td>
<td>35</td>
</tr>
<tr>
<td>Lattice Contacts</td>
<td>36</td>
</tr>
<tr>
<td>A15G with 3’U Overhang</td>
<td>37</td>
</tr>
<tr>
<td>C38G and A59G with 3’U Overhang</td>
<td>39</td>
</tr>
<tr>
<td>C7A with 3’U Overhang</td>
<td>41</td>
</tr>
<tr>
<td>91-add-GU Plus A15G and 3’U Overhang</td>
<td>43</td>
</tr>
<tr>
<td>Molecular Modeling</td>
<td>45</td>
</tr>
<tr>
<td>Relationship between Sub-domains in the SAM-I Family</td>
<td>48</td>
</tr>
<tr>
<td>Isothermal Titration Calorimetry</td>
<td></td>
</tr>
<tr>
<td>TOPIC</td>
<td>PAGE NUMBER</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
</tr>
<tr>
<td>CONCLUSIONS</td>
<td>53</td>
</tr>
<tr>
<td>APPENDIX A</td>
<td>56</td>
</tr>
<tr>
<td>REFERENCES</td>
<td>58</td>
</tr>
</tbody>
</table>
Acknowledgements

I would like to thank all the members, past and present, of the Batey Laboratory for their help and comradery throughout the time spent on my research. Because of them, I have grown intellectually as a student and researcher. I would like to give a special thanks to Dr. Francis Reyes, who first assisted me and helped me with the biochemical experiments and techniques used in the laboratory, as well as Dr. Andrea Edwards, whose groundwork investigation on the SAM-I/IV gave the foundation for my independent research project. In addition, without the patience, assistance, and contributions from Jeremy Trausch, this study would not have been possible. I would also like to express my immense gratitude to Dr. Robert Batey, who provided me with the opportunity to carry out independent research within his laboratory. His vast knowledge in the field as well as great insight into my project and thesis has undoubtedly enriched my undergraduate experience. I am grateful for the Undergraduate Research Opportunities Program as well as the Howard Hughes Medical Institute for providing funding throughout the duration of my research. Finally, I would like to thank my parents for providing support every step of the way. The work presented here could not have been done without the assistance and support of all mentioned above.
Abstract

Riboswitches are elements that directly bind small molecule metabolites without the assistance of other factors. Riboswitches are found in the leader sequence of bacterial as well as some plant and fungi mRNA’s. They are highly specific towards the metabolite that they bind which helps them play a significant role in gene regulation. The S-adenosylmethionine (SAM) sensing riboswitches modulates the expression of genes involved in the sulfur metabolism and methionine biosynthesis. One family of SAM sensing riboswitches that has been briefly studied by the Batey laboratory is the SAM-I/IV riboswitch. A preliminary structure was solved at 3.5 Å resolution. While this resolution was sufficient to generate a large amount of the model, a higher resolution image of the riboswitch is essential for visualizing the details of the AD-ligand complex. A high resolution image of this region is necessary in order to clearly see the binding mechanism of this complex, which will help us fully understand the directing structure-function studies on this RNA. This study was performed with the intent to find optimal crystallization conditions as well as to implement point mutations within the crystal construct to improve the resolution of the SAM-I/IV riboswitch structure. A molecular model was made that helped clarify important regions in the wild-type structure.
Introduction

Gene Regulation

Gene regulation is paramount for cell survival. Regulation of genes expressed in the cell need to be controlled to maintain the appropriate levels of RNA and proteins in order to maintain vital cellular processes (Maumita and Breaker, 2004). Many factors within the cell need to work together in order to quickly and efficiently interrogate the cellular environment in order to execute a response. Vast amounts of energy are expended in the cell during these intricate and elegant interactions to maintain the vital cellular processes. Even the simplest single-cellular organisms need to modulate the expression of hundreds of genes in response to a multitude of cellular needs and environmental signals (Maumita and Breaker, 2004).

Proteins were once thought to exclusively control regulation of genes through changes in the intracellular environment through their role as activators, repressors, initiation factors, and feedback inhibitors (Winkler and Breaker, 2003). Examples of this can be seen in processes such as translation, transcription, mRNA processing/degradation as well as other various mechanisms.

One of the greatest insights in the post genomic era is the extent in which RNA participates in gene regulation. Current discoveries that showcase novel RNA regulatory pathways accentuate how our working knowledge and appreciation of cellular regulation is vastly deficient. A recent study found only one-fifth of transcription across the human genome is associated with protein-coding genes (Kapranov, 2007). This indicates that there are at least four-times more long non-coding than coding RNA sequences (Kapranov, 2007). If this is the case,
then perhaps there are greater and more intricate regulation processes that are essential to cell survival.

Charles Yanofsky’s groundbreaking discovery of transcriptional attenuation of the tryptophan operon in *Bacillus subtilis* helps show the intricate role RNA plays in gene regulation (Figure 1). It was found that proteins involved in tryptophan synthesis were controlled by the tryptophan (*trp*) operon (Du *et al*., 2000). The transcription of this operon is controlled by *trp* RNA-binding attenuation protein (TRAP) (Du *et al*., 2000). In the absence of *trp*, the protein is inactive and the *trp* leader sequence forms a stable anti-terminator structure (Du *et al*., 2000). When TRAP is activated by high levels of intracellular *trp*, it binds to an RNA target sequence, which overlaps with the anti-terminator structure and prevents its formation (Du *et al*., 2000). Instead, an overlapping terminator structure forms and halts the transcription (Du *et al*., 2000).

![Figure 1. Transcription attenuation model of the *B. subtilis trp* operon. (A) Mechanism of attenuation during conditions of tryptophan excess. (B) Formation of anti-terminator during limiting tryptophan conditions. TRAP is represented by the gray doughnut structure. Adapted from (Du *et al*., 2000).](image-url)
Another example of RNA mediated gene regulation can be seen with rho-independent transcriptional regulation in which the RNA will form mutually exclusive structures that will allow the RNA polymerase to proceed with the read through or abort the synthesis of the message altogether (Batey, 2006). Formation of these distinct secondary structures depends on interaction of protein (often in conjunction with a small cellular metabolite) with the leader sequence of the mRNA (Batey, 2006).

It is well established that RNA-protein interactions can directly sense biological signals and modulate gene expression. However, regulatory elements in the leader sequence of RNA’s have been recently shown to display highly selective binding affinity of small-molecule metabolites without the aid of proteins (Sudarsan et al., 2003). These elements, dubbed “riboswitches”, bind certain metabolites with high affinity in a cis fashion to control transcription or translation (Sashital and Butcher, 2006).

Riboswitches

Riboswitches are unique control elements primarily found in the leader sequence of messenger RNAs (mRNAs) in prokaryotes as well as in plant and fungal introns and 3’-UTRs (Sudarsan et al., 2003). These natural biosensors selectively bind metabolites, which in turn induce a change in the mRNA structure to control gene regulation. While classical genetic studies provided evidence for the existence of the riboswitch as early as the 1970’s (Escalante-Semerana and Roth, 1987), they were not formally proved until 2002 by Nahvi et al. when the authors provided evidence for a protein-free binding event between the leader sequence of the
*E. coli* *btuB* I mRNA and coenzyme B$_{12}$ (Nahvi et al., 2002). This binding event established a unique RNA structure and subsequent repression of BtuB protein expression (Nahvi et al., 2002). In addition, the *RFN* element, which frequently is found in the 5'-untranslated region of genes responsible for the biosynthesis or import of riboflavin and FMN, serves as the receptor portion of FMN-dependent riboswitches in *Bacillus subtilis* (Sudarsan *et al*., 2003).

So far, riboswitches have been identified in the leader sequence of mRNAs throughout many different organism including both Gram negative and Gram positive bacteria (Nudler and Mironov). In addition, riboswitches have been found in plants, including *Arabidopsis thaliana*, as well as in fungi, including *Neurospora crassa* and *Fusarium oxysporum* (Sudarsan *et al*., 2003). Sequence alignments of riboswitches found in both prokaryotes and eukaryotes show conserved nucleotide positions within their homologous domains (Sudarsan *et al*., 2003).

**Structure**

Riboswitches are conceptually divided into two parts: the aptamer domain (AD) and the expression platform (EP). The AD folds into a highly complex three-dimensional structure that creates the ligand binding pocket. Riboswitches have been shown to bind a wide variety of metabolites. Some metabolites are simple in structure, such as glycine, to very complex such as vitamin B$_{12}$ (Figure 2). The secondary structures within the AD are typically highly conserved and unique to their class (Winkler, Cohen-Chalamish and Breaker, 2002). The AD takes the place of the conventional protein cofactor that would otherwise serve as the sensory element (Batey, 2006).
In all cases, the AD is highly specific for a single small molecule recognizing even the slightest variations among atom or functional group placements (Winkler and Breaker, 2002). One case of this is illustrated by the FMN riboswitch. This riboswitch binds FMN with 1,000-fold higher affinity than riboflavin, differing only in the placement of a single phosphate residue (Mandal and Breaker, 2004). The S-adenosylmethionine binding riboswitch (SAM-I) binds a metabolite analog, S-adenosylhomocysteine (SAH) with 1000 fold decreased affinity (Epshtein et. al., 2005). Likewise, the Purine riboswitch binds guanine with 1000 fold affinity for adenine due to a Watson-Crick base pair with the ligand (Gilbert et. al., 2009).
The EP is typically located downstream of the AD. The role of the EP is to interpret the signal sent by the AD based upon binding interactions with the ligand. This signal will then instigate a conformational change in the leader sequence (Batey, 2011). In riboswitches that attenuate transcription, the two mutually exclusive secondary structures of the EP are a rho-independent terminator that causes RNA polymerase to disengage, or an anti-terminator that allows transcription to proceed and synthesize the entire message (Figure 3).

![Figure 3](image-url)  

**Figure 3.** An Example of the Riboswitch Mechanism. In the absence of ligand (L), an intrinsic terminator stops transcription. Upon ligand binding, an anti-terminator helix is formed allowing transcription of the genetic message. Adapted from (Topp and Gallivan, 2010).

This switching mechanism relies on a “switching sequence” common to both domains (Edwards and Batey, 2010). Placement of the switching sequence in the aptamer domain or the expression platform ultimately dictates the expression outcome of the mRNA (Edwards and Batey, 2010) (Figure 4). Riboswitches that control transcription utilize the switching sequence to direct the EP to form a rho-independent transcriptional terminator, a short stem-loop that signals for RNA polymerase to abort transcription (off-state) or an anti-terminator which allows synthesis of the entire message (on-state) (Edwards and Batey, 2010). Other riboswitches can...
control regulation at the translational level, in which case the switching sequence is employed to expose or sequester the Shine-Dalgarno sequence (Edwards and Batey, 2010).

A simple example can be seen in the mechanism of riboswitch SAM-I, which regulates gene expression at the transcriptional level. Binding of S-adenosylmethionine to the AD induces a terminator stem loop to stabilize in the EP, which in turn causes early termination of transcription (Mironov et al., 2002) (Figure 5).
Application

The complex structures, binding capabilities, and gene regulation of riboswitches make them a target for ongoing research to understand and characterize these properties. Furthermore, a complete understanding of how riboswitches regulate essential cellular processes is a priority. Advancements in high-throughput sequencing and genomics have resulted in new insights regarding the important relationships between the human microbiome and health (Pflughoft, et.al., 2011; Zhu et al., 2010). It has been found that the ecology of the microbiome in the human gut is significantly different amongst obese individuals and individuals of normal weight (Turnbaugh et al., 2006). Other studies suggest that the human microbiome plays an important role in a number of disease processes (Pflughoft, et.al., 2011; Zhu, et.al., 2010). In the future,
being able to control the pathways in this particular bacterial ecology may become an important avenue in medical science, which would help give ample insight into human health.

Riboswitches could also be targets for novel antibiotics. Indeed, some antibiotics whose mechanism of action was unknown for decades have been shown to operate by targeting riboswitches. For example, when the antibiotic pyrithiamine enters the cell, it is metabolized into pyrithiamine pyrophosphate. Pyrithiamine pyrophosphate has been shown to bind and activate the TPP riboswitch, causing the cell to cease the synthesis and import of TPP. Because pyrithiamine pyrophosphate does not substitute for TPP as a coenzyme, the cell dies (Blount and Breaker, 2006).

Another example of using riboswitches as targets for antibiotics is illustrated by the mononucleotide (FMN) dependent riboswitch (Lee et al., 2009). This riboswitch binds FMN in order to regulate the expression of genes responsible for the transport and biosynthesis of riboflavin, a precursor of FMN (Lee et al., 2009). Roseoflavin, a chemical analog of riboflavin and FMN that also exhibits antimicrobial properties, can also bind to the FMN riboswitch (Lee et al., 2009). It was observed that upon riboswitch binding to roseoflavin in L. monocytogenes, expression of the gene located downstream of the FMN riboswitch, a riboflavin transporter, was down-regulated (Mansjo and Johansson, 2011). As a result, growth of the bacterial cell was inhibited (Mansjo and Johansson, 2011).

Furthermore, there has been interest in application of riboswitches as biotechnological tools (Topp and Gallivan, 2010). Ligand-inducible expression systems have served as important
genetic tools to increase protein to reasonable amounts for biochemical study (Topp and Gallivan, 2010). However, these tools can be quite expensive, and using natural riboswitches activated by amino acids may prove to be an effective alternative (Topp and Gallivan, 2010). It has been reported that the glycine riboswitch from *Bacillus subtilis* was used for glycine inducible production of β-galactosidase in *Bacillus subtilis* cells (Phan and Schumann, 2007). Although this system only produced six-fold induction, the result showed promise if the strategy can be strengthened where more ideal induction parameters can be implemented (Topp and Gallivan, 2010).

The ability to modulate bacterial motility in response to arbitrary chemical signals would provide new tools for bioremediation and drug delivery (Topp and Gallivan, 2010). It was shown that the *E. coli* chemotaxis system could be re-programmed by placing a key chemotaxis signaling protein (*cheZ*) under the control of a theophylline-sensitive riboswitch (Topp and Gallivan, 2010). In doing so, *E.coli* cells could distinguish between a path of theophylline and one of caffeine and allowed it to selectively follow the path of theophylline (Topp and Gallivan, 2010) (Figure 6). This precise localization to the ligand represents a sharp contrast to the behavior of wild type *E. coli* (which do not stop moving, and thus cannot localize to a chemical signal), and such behavior may prove useful for targeting cells (Topp and Gallivan, 2010).
One of the first riboswitches to be identified was the SAM-I riboswitch (Figure 7). This riboswitch is prevalent in Gram positive bacteria and acts to help regulate methionine biosynthesis, SAM biosynthesis, and sulfur metabolism within the cell. SAM is the second most abundant intracellular metabolite and serves as a primary methyl donor for many cellular processes. The riboswitch mechanism relies on tight binding of SAM to the aptamer, partitioning the switching sequence into the AD to stabilize the formation of a terminator stem loop to promote early termination of transcription (McDaniel, 2003).

Figure 6. Caffeine (red) Tryptone (blue) and Theophylline (green) are shown to be plated. *E. coli* cells containing the *cheZ* gene under the control of a theophylline-sensitive synthetic riboswitch are plated at the bottom. The cells can distinguish between the compounds, and will exclusively follow the path of Theophylline, as seen by the figure on the right. Figure adapted from (Topp and Gallivan, 2010).

*S-adenosylmethionine Binding Riboswitches*

Figure 7. Cartoon structure of SAM-I solved by the Batey Laboratory in 2006. SAM metabolite (purple) is shown to be bound.
Studies have shown that a large variety of transcriptional units served as targets for SAM (Winkler and Breaker, 2003). Further investigations of these transcriptional units have led to the discovery of seven unique riboswitch families that are targets for SAM. These families are: SAM-I, SAM-II, SAM-III, SAM-IV, SAM-V, SAM/SAH, and SAM-I/IV. Crystal structures for three of the seven known families of SAM riboswitches have been solved by the Batey laboratory (Montange and Batey, 2006; Batey et al., 2008; Batey et al., 2010).

While these families bind SAM, not all families share a conserved binding core sequence. Interestingly, three of these families appear to share the same core, but have very different three-dimensional architectures (de la Pena et al., 2009). For example, it has been observed that the SAM-I, SAM-IV, and SAM-I/IV families share a binding core that is nearly identical in sequence, yet very different from the binding core of the SAM-II, SAM-V, SAM/SAH, and SAM-III families. Similarly, SAM-II, SAM-V, and SAM/SAH have similar binding cores that are different from the other families. Further review can be found in (Poiata et al., 2009; Edwards and Batey, 2010). By investigating these differences, it will give us a better understanding of the structural basis for the relationship of SAM-I/IV to SAM-I.

*SAM-I Riboswitch*

Recognition of the binding pocket in SAM-I allows the mobile methionine moiety to fold over the adenine ring in which the cation from the nitrogen atom is stabilized through interaction with the pi-system of the adenine ring (Montange, 2006). Once in the cis-conformation, both the methionine main chain and adenine ring interact with the binding pocket (Montange and Batey,
The adenine ring of SAM is involved in a base triple with A45 and U57 where U57 pairs with the SAM adenine ring in a Hoogsteen fashion (Montange and Batey, 2006). The main chain atoms in the methionine moiety are recognized through the formation of a quartet with G11, C44 and G58, where G11, C44 and G58 form a base triple and the carboxylate of the methionine main chain hydrogen bonds with the Watson-Crick face of G11 (Montange and Batey, 2006). The amino group is recognized by bonding with the ribose ring of G58 (Figure 8).

Figure 8. a) Recognition of the adenosyl moiety of SAM by an internal loop with P3. b) Recognition of the methionine moiety by elements of P3 and J1/2 (G11). c) Recognition of the positively charged sulfonium moiety by two universally conserved base pairs in P1.
Similarities between SAM-I and SAM-IV Riboswitches

It has been thought that the SAM-I and SAM-IV families share similar cores, yet support vastly different peripheral moieties which allows them to take on different three dimensional architectures. Therefore, the core of the SAM-I and SAM-IV riboswitches might share similar binding sites and molecular recognition characteristics, yet allow for different global architectures to be supported (Weinberg et al., 2008). This hypothesis is supported by the SAM-I atomic-resolution model in which 6 nucleotides within the core were proposed to interact directly with the ligand (Figure 9). Of the six positions in SAM-IV proposed to correspond to these SAM-I ligand-contacting positions, five are conserved in all known SAM-IV representatives with the same nucleotide identity as in SAM-I riboswitches (Weinberg et al., 2008). The only nucleotide difference is that uracil in position 88 (U88) in SAM-I is occupied by a cytosine in SAM-IV.

Figure 9. Secondary structures of SAM-I and SAM-IV. The common nucleotides common to both riboswitches are highlighted in yellow. Peripheral structures that are unique to both are also depicted. The novel pseudoknot in SAM-IV between P3 and P5 can be seen as well. Adapted from (Weinberg et al., 2008).
**SAM-I/IV Riboswitch**

One of these families of SAM riboswitches that has been investigated briefly by the Batey lab is the “SAM-I/IV” family of riboswitches. The core of SAM-I/IV displays similarity to the cores of two other riboswitches, the SAM-I and SAM-IV families (Weinberg *et al.*, 2010). Because the core of SAM-I/IV is very similar to that of SAM-I and SAM-IV, it has been hypothesized that molecular recognition of SAM to the binding pocket of SAM-I/IV would be similar to that of both SAM-I and SAM-IV. However, differences in the overall architectures amongst the three riboswitches may lead to differences in the switching events between the aptamer domain and expression platform. While both SAM-I and SAM-IV share pseudoknot motifs, SAM-IV forms an additional pseudoknot between P3 and P5, which is vastly different than the pseudoknot formed in SAM-I (Figure 10).

![Figure 10](image.png)

**Figure 10.** The SAM-I Clan. Schematics for SAM-I, SAM-IV, and SAM-I/IV are shown above to show similarities between the binding pockets and differences among the architectural peripheries. SAM ligand is represented by the red and yellow oval labeled “SAM.” PK-1 can be seen in the region colored green, which can be found in SAM-I and SAM-IV. PK-2 is highlighted in the cyan color region, which can be found in SAM-IV and SAM-I/IV.
This second pseudoknot allows SAM-IV to regulate gene expression differently than SAM-I. Whereas binding of SAM to the binding pocket of SAM-I destabilizes the anti-terminator helix, binding of SAM to the binding pocket of SAM-IV destabilizes a terminator stem-loop (Weinberg et al, 2008).

While a formal structure of the SAM-IV riboswitch has not been solved, nor has a complete switching mechanism been resolved, it is thought that P5 forms a stem loop in which the Shine-Dalgarno sequence is sequestered, preventing the 30S subunit of the ribosome to bind (Weinberg et al, 2008). Once SAM occupies the binding pocket of the riboswitch, the terminator stem loop is melted. The single strand of RNA is stabilized upon formation of the pseudoknot with P3 (Weinberg et al, 2008). This unveils the Shine-Dalgarno sequence upstream of the AUG start codon allowing the genetic message to be read. The PK-2 motif in SAM-I/IV is similar to the pseudoknot formed between P3 and the 5’-side of P5 in SAM-IV (Figure 8). Because of this, it is thought that similarities between these two pseudoknot formations would lead to similar mechanisms in gene regulation.

Research Objective

To understand the structural basis of the specific recognition of SAM and how binding is transduced into changes in gene expression in SAM-I/IV, the Batey lab has solved the structure of this riboswitch at 3.5 Å resolution (Edwards and Batey, unpublished data). The structure solved at 3.5 Å resolution allowed for a preliminary understanding for the structural basis of SAM binding. While this resolution was sufficient to generate a large amount of the model, a
higher resolution image of the riboswitch is essential for visualizing the details of the AD-ligand complex as well as the stability of the novel PK-2 element. As of now, the image of the AD-ligand complex is not clear and prohibits definition of the binding regions (Edwards and Batey, unpublished data). A high resolution image of this region is necessary in order to clearly see the binding mechanism of this complex, which will help us fully understand the directing structure-function studies on this RNA. Crystallization reproducibility is also an issue. It was determined that the initial crystals which diffracted to 3.5 Å could not be replicated. In order to collect data to solve a structure at a higher resolution, a protocol that gives reproducible results is critical.

The major scope of this project will be to address these current issues at hand. The preliminary crystallization conditions for the native RNA will be used as a starting point from which optimizing strategies will be implemented. One technique that will be used that has worked in optimizing RNA lattice quality is by introducing point mutations into areas of the RNA that are believed to exhibit structural strain (Reyes, et.al, 2009). These point mutations will not only help stabilize this area in order to create a better defined lattice structure which will produce a clear image of the riboswitch, but will also help with crystallization reproducibility (Edwards, et.al, 2010). Investigations of our preliminary structure suggest the additional nucleotides at the 3’end of the RNA may facilitate crystal formation.

Secondary structures across the SAM-I/IV families were analyzed in order to assess which nucleotides could be changed to exploit more favorable intra and inter-molecular base interactions while preserving the conserved nucleotides that were deemed crucial for SAM binding and/or architectural stability (Figure 11). From this analysis of base conservation
patterns of the SAM-I/IV family, we implemented further strategies to change certain nucleotides within the backbone of the RNA in hopes to increase intermolecular lattice contact stability. Two additional aptamers will be created to incorporate the peripheral elements of both the SAM-I and SAM-I/IV riboswitch, mainly showcasing the two main pseudoknots.

**Figure 11.** Basic secondary structure of the SAM-I/IV riboswitch that shows the nucleotides that are conserved as well as the ones that are available to change. Arrows show sites for future nucleotide mutations that could be sites that exhibit structural strain or form potential lattice contacts. The legend denotes the percentage of conservation among nucleotide identities. Adapted from (Breaker, 2011).

Resolution of these problems will help provide insight into a more detailed binding mechanism, as well as providing insight with regards to see if the peripheries of the SAM-I/IV and SAM-I riboswitch are mutually exclusive.
Methods

DNA Design/Synthesis

The SAM-I/IV riboswitch was created from the env87 variant found within a Pacific Ocean Metagenome (accession number ABEF01012528.1). The wild type sequence and all mutants thereof into a DNA transcription template were introduced by developing oligonucleotides. Complete DNA fragments were constructed by using 5’ GEN as the 5’ primer followed by addition of A, B, C, and D to be used as complements for extension in each mutant construct (Appendix A). Water was added to the primers such that each primer had a final concentration of 100 mM. With this, a PCR protocol was followed, making sure to add PFU DNA polymerase last (Table 1). The PCR reaction was done for 30 cycles of 95 °C, for 30 seconds, 55 °C for 30 seconds, 72 °C for 1 minute, and a hold at 16 °C. After the program had finished, the DNA was checked for quality using a 1% agarose gel against a 1000kb+ DNA ladder to check for correct nucleotide length (Figure 12).

Table 1. Reagents and Volumes for PCR construction.

<table>
<thead>
<tr>
<th>Reagents</th>
<th>Volume (μL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddH2O</td>
<td>840</td>
</tr>
<tr>
<td>10x PFU Buffer</td>
<td>100</td>
</tr>
<tr>
<td>1:25 PFU enzyme</td>
<td>10</td>
</tr>
<tr>
<td>10mM dNTP</td>
<td>20</td>
</tr>
<tr>
<td>100μM 5’Oligo</td>
<td>10</td>
</tr>
<tr>
<td>100μM 3’ Oligo</td>
<td>10</td>
</tr>
<tr>
<td>1:100 dilution of inner oligos</td>
<td>10</td>
</tr>
</tbody>
</table>

*PFU buffer consisted of 200mM Tris-HCl (pH 8.8), 100mM (NH4)2SO4, 100mM KCl, 1% (v/v) Triton X-100, 1mg/ml BSA and 20mM MgSO4.
RNA Synthesis

After the DNA was constructed and checked to ensure a correct length, the DNA was transcribed *in vitro* using the following protocol: To a 50 milliliter conical tube, the PCR product was added along with other necessary transcription reagents (Table 2). The reaction mixture was then incubated in a 37 °C bath for four hours. Absolute ethanol was added to a final volume of 50 milliliters to precipitate the RNA. The mixture was stored at -20 °C overnight.

### Table 2. Reagents and Volumes for RNA transcription.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume (mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ddH&lt;sub&gt;2&lt;/sub&gt;O</td>
<td>7</td>
</tr>
<tr>
<td>10x Transcription Buffer*</td>
<td>1.25</td>
</tr>
<tr>
<td>1M MgCl&lt;sub&gt;2&lt;/sub&gt;</td>
<td>0.4</td>
</tr>
<tr>
<td>1M DTT</td>
<td>0.1</td>
</tr>
<tr>
<td>100mM rNTP</td>
<td>0.5 x 4 (A,G,C,U)</td>
</tr>
<tr>
<td>PCR Template</td>
<td>1</td>
</tr>
<tr>
<td>T7 RNA Polymerase</td>
<td>50 µL</td>
</tr>
<tr>
<td>IPPase**</td>
<td>200 µL</td>
</tr>
</tbody>
</table>

*10x Transcription buffer consisted of 15 mL 1 M Tris pH 8.0, 5 mL 1 M DTT, 0.25 g Spermine, 50 µL Triton X-100 filled to 50 mL of ddH<sub>2</sub>O.

**IPPase stands for Inorganic Pyrophosphatase which is used to catalyze the reaction between one molecule of inorganic phosphate into two phosphate ions which serves as the driving force for nucleotide addition during primer extension.

![Figure 12. 3’U overhang DNA construct (lane 2) run against a 1000 kb+ ladder (lane 1) on a 1% agarose gel. Base Pair (BP) marker lengths are indicated with red arrows.](image)
**RNA Purification**

The ethanol precipitated RNA was recovered by centrifugation at 3000 rpm for 20 minutes. The ethanol was discarded and the remaining pellet was resuspended in two milliliters of 8M urea, 750 µL of 0.5x TE, which is 10 mM Tris (trishydroxymethylaminomethane) brought to pH 8 added to 1 mM EDTA, and 1500 µL of formamide loading dye.

The prepared RNA was loaded into an 8% denaturing acrylamide slab gel containing 8 M urea and ran at 35 W. Afterwards, the gel was taken out and examined under ultra-violet light. The largest band of interest was marked and cut out of the gel. The cut band was then placed into another 50 milliliter conical and 0.5x TE was added to the 50 milliliter line. The conical was then placed on a shaker and mixed for at least six hours in order for the RNA to diffuse out of the gel. Next, the conical was centrifuged for 20 minutes in order to separate the solid gel from the RNA solution. The RNA was concentrated using a 10 kD molecular weight cut-off centrifugal concentrator and then buffer exchanged with 0.5x TE three times. The concentration of the RNA was then measured using a spectrophotometer and recording the absorbance value at 260 nm. An extinction coefficient was determined using an online extinction coefficient calculator. Using Beer’s Law ($A=\varepsilon cl$) and the extinction coefficient determined for the RNA of interest, the concentration of the RNA could then be determined. The RNA was then loaded onto an 8% denaturing acrylamide gel to check for purity (Figure 13).
Crystallization Condition Screen

24-well Nextal trays were used to screen a variety of mother liquors by use of hanging drop vapor diffusion. An RNA solution consisting of a final concentration of 2 mM SAM, 5 mM MgCl₂, and 300 μM RNA was made. All necessary mother liquor reagents were added to the tray wells to a final volume of 500 μL in the well. Exact crystallization conditions for each construct can be found in the results section. 1.5 μL of RNA solution was added to 1.5 μL of mother liquor on a glass cover slip. The slip was sealed over the well and the final tray was allowed to incubate at 30 °C. Crystal growth was observed 48 hours after tray set up.

A screening process for optimal crystal growth was then implemented. Crystals were scored on a scale of 1-10 based on qualitative judgments that categorized crystals based on size, morphology, and distinctness of edges. A crystal that was scored a 10 had a large size (about 100 μm), and had distinct edges (meaning that the shape and morphology were clearly defined) whereas a crystal that was given a 1 usually was small (less than 10 μm and too difficult to scoop), and lacked well defined edges. While a crystal that produces a quality diffraction pattern

Figure 13. Two samples of purified add-91-GU with A15G and 3′U overhang RNA constructs (lanes 4 and 5) ran against un-purified 3′ U overhang constructs used as markers (lanes 1, 2 and 3) are shown above.
does not necessarily need to be scored highly, it has been observed that there is a strong correlation between diffraction quality and crystal size and definition of crystal edges (Guinier, 1994). Once the crystals were scored from the trays, a new grid was made that incorporated the conditions of the best crystals that were produced. The method was repeated until high quality crystals could be reproducibly grown.

**X-Ray Diffraction**

To prepare crystals for diffraction analysis, they were first cryoprotected in order to protect the crystal from “freeze damage,” which is due to the formation of ice during freezing. A cryoprotecting protocol was to add 20 μL of MPD diluted to 20% in original mother liquor directly on top of the drop on the glass cover slip. The crystals were allowed to soak for three minutes and then mounted in a nylon loop. Once the crystals were removed from the drop, they were frozen in liquid nitrogen and then imaged on a Rigaku MSC X-ray diffractometer with a copper anode. A preliminary image screen was performed to assess if the crystal was of high enough quality to collect a full diffraction image set. The preliminary screen protocol used was to image the crystal using five minute beam exposures with rotations at 0, 30, 60, and 90 degrees. Crystals were also sent to the Advanced Light Source Synchrotron in Berkeley, California in hopes to obtain high resolution data.
Molecular Modeling

As stated previously, a 3.5 Å resolution model of the riboswitch was solved. Because the structure solved did not include G91 or U92 in the J5/PK2 region, it was deemed necessary to include these nucleotides back in to give a more complete and correct structure of the riboswitch. A molecular model was constructed that included G91 and U92 back into the J5/PK2 region. It was observed that these additional insertions had significant structural changes in the pseudoknot. Explanations concerning how the model was constructed will be further detailed in the Molecular Modeling subheading in the results and discussion. Energy minimization techniques were used to optimize geometry of the RNA to allow superimposed strands to fit perfectly. Optimization of bonding angles/lengths, hydrogen bonding, and electrostatic interactions were optimized while steric clashing and unfavorable Van der Waals interactions were minimized using the program, CNS 1.3. This work was primarily done by my advisor, Dr. Robert Batey.

Isothermal Titration Calorimetry (ITC)

ITC is a thermodynamic technique used to quantitatively characterize bimolecular interactions with high precision and accuracy (MicroCal, 2003). With this machine, the pipette will titrate ligand into the sample cell filled with RNA solution (Figure 14). When ligand binds to riboswitch, there will be a structural change in the riboswitch’s tertiary structure as it changes from its unbound form to bound form. The heat of reaction can be experimentally detected and its data can be extrapolated to give insight into the binding mechanism of ligand to riboswitch. The data can then give precise measurements of dissociation constant (K_d), the stoichiometric...
ratio (n) of the binding molecules, and the thermodynamics of binding (changes in enthalpy $\Delta H$, entropy $\Delta S$, and free energy $\Delta G$) (MicroCal, 2003).

![Diagram of an ITC machine.](image)

**Figure 14.** Diagram of an ITC machine.

Concentration of the titrant (ligand) in solution is usually 10 fold higher than concentration of the substrate (Batey and Gilbert, 2009). ITC data collected from each experiment was fitted to a single-site binding model using Origin ITC software (Microcal Software Inc), from which values of $n$, $K_a$ (M$^{-1}$), $\Delta H$ (cal/mol), and $\Delta S$ (cal/mol*K) were extrapolated using the equation:

$$q = (v) \left( \Delta H \right) ([RNA]) \left( (K_a[L]^n/1 + K_a[L]^n_i) - (K_a[L]^{n-1}_i/1 + K_a[L]^{n-1}_i) \right)$$
where \( q \) is the heat released, \( v \) is the known volume of the reaction, \( K_a \) is the association constant, and \( L_i \) is the ligand concentration at the \( i^{th} \) injection (Leavitt & Freire, 2001). The disassociation constant, \( K_d \) (M), was calculated by taking the inverse of \( K_a \). In order to validate that the changes made to SAM-I/IV (\( \Delta G_91, U92 \)) are compatible to ligand binding, the affinity for SAM among the SAM-I/IV crystal construct, SAM-I/IV GU insertion, and the wild type RNA were measured.

**ITC Sample Preparation**

Purified RNA was placed into 6-8 kD MWCO dialysis tubing. The RNA solution was split and dialyzed overnight against one liter of buffer consisting of 135 mM KCl, 15 mM NaCl, 10 mM Hepes, pH 8, and 1 mM MgCl\(_2\), as well as one liter of buffer containing 135 mM KCl, 15 mM NaCl, 10 mM Hepes, pH 8, and 10 mM MgCl\(_2\). SAM was prepared through the dilution of a 100 mM stock using the same buffer used for dialysis of the RNA to ensure an exact buffer match and to avoid measurement of heat of dilution instead of binding events (Wise, 2004). The SAM solution was then accurately measured for concentration with the spectrophotometer at 260 nm using the extinction coefficient of 15400 M\(^{-1}\) cm\(^{-1}\). All ITC experimental parameters were set at 37 °C, a reference power of 11 µcal/sec, an initial delay of 60 seconds, and a titration of 21 injections each of 1.6 µL with a 3.2 second duration spaced 180 seconds apart. Each experiment was carried out in triplicate.
Results and Discussion

Structural Analysis of SAM-I/IV

In order to address issues concerning the clarity of the image of the AD-ligand complex and its function on how it effects formation of the three-dimensional structure of the riboswitch during ligand binding, a higher resolution image will be needed. In order to generate this image required for a detailed structural analysis, a search for optimal mother-liquor conditions that allow reproducible crystals of high diffraction quality to be grown will be done. In addition, mutations within the sequence of the RNA that might promote favorable stabilizing interactions will be implemented in hopes to increase image resolution.

Crystallography of SAM-I/IV Mutants

The original structure showed evidence for a non-templated uridine residue at the 3’end that appeared to form a critical stabilizing contact with A33. Further, it was discovered that there was a GU deletion in a region that might impair formation of PK2. Together, these issues might explain the irreproducibility of the crystals produced by Dr. Andrea Edwards.

Strategies implemented to try and resolve these issues were to introduce point mutations into the RNA at sites likely to mediate lattice contacts or sites that are poorly structured in the RNA, such as single bulged nucleotides (Reyes, Garst, and Batey, 2009). These point mutations will help stabilize these regions through base paring or base stacking interactions. Base pairing occurs when nucleotides form hydrogen bonds between one another, which can aid in the stability of the helical structure that is hallmark of DNA and RNA. Base stacking is also an
important stabilizing interaction between nucleotides that help form the helical shape as well as to help provide stable lattice contacts. The driving force of this stability is due to overlap of the pi-electrons between nucleotides due to their conjugated pi-system. Overlap of this pi-cloud between bases also helps exclude water molecules that might otherwise form a clathrate (a cage like structure) around the hydrophobic base. Release of these water molecules increases overall entropy, helping stabilize the base stacking interaction. In addition, the polarizable pi-cloud of the nucleobases induces parallel dipoles that attract each other, further stabilizing the base stacking interaction. Bases with multiple aromatic rings contribute to these interactions even more. Base stacking among purines, and in particular guanines, are highly favored. Because of this, strategies to create GG base stacks were implemented.

Point mutations at the 3’ end of the RNA as well as nucleotide alterations in the J5/PK2 region were made in order to try and optimize the stability of the pseudoknot through use of base pairing and base stacking. It was initially thought that reducing the strain in this region would also help overall stability of the RNA molecule. By doing so, perhaps the change in structure of the pseudoknot to a less strained confirmation would help promote favorable lattice contacts in this region, and thus give diffraction at a higher resolution. The mutant 3’ overhangs made were the 3’U, 3’UC, and 3’G overhang.
"3’U Overhang"

The first preliminary structure solved suggested there to be a single non-template uracil overhang at the 3’ end of the RNA. It was determined that the uracil overhang formed a Watson-Crick base pair with A33 in the J5/PK2 region, helping stabilize the formation of the pseudoknot. This stability was thought to help the molecules pack together more securely within the crystal lattice structure. A screening process was done as described in the methods section and it was found that large crystals of a long cylindrical morphology could be reproducibly grown in the conditions of 0.05M Na-Cacodylate pH 8 (Na-Cac), 0.04 M MgOAc, 0.30M KCl, 7% MPD, 0.001 M Cobalt Hexamine (CoHex). 20 μL of cryoprotectant, which consisted of 2-Methyl-2,4-pentanediol (MPD) diluted to 20% in tray well mother liquor, was added on top of the drop and allowed to soak for three minutes. Afterwards, the crystals were analyzed using X-Ray diffraction. While these crystals were of diffraction quality, they did not however surpass the resolution of 2.5 Å, which was the intended goal. The highest resolution image obtained was 3.7 Å (Figure 15). Even though the 3’U overhang was able to stabilize the pseudoknot enough for crystals to be reproduced, perhaps there was still significant structural strain in this region that was preventing optimal crystal lattice contacts from forming.
Figure 15. a) Secondary structure of SAM-I/IV depicting the 3' U overhang boxed in blue. b) Crystal obtained of SAM-I/IV. c) Diffraction pattern of crystal (b).
3’UC, 3’G Overhang, and no 3’U Overhang

3’UC and 3’ G overhangs were added to try and improve stability of the pseudoknot through base pairing and base stacking interactions (Figure 16 a and b). While a 3’ G overhang would not form a Watson-Crick base pair with A33 in the J5/PK2 region, it was thought that perhaps the flexibility of the 3’ overhang would allow for a favorable base stacking interaction between the 3’ G overhang and A33. The 3’ UC overhang was thought to not only form a Watson-Crick base pair between U96 and A33, but perhaps a base stacking interaction could be formed between C97 and G32 which would provide additional stability to this region. A control was also made that lacked a 3’ overhang overall in order to more closely resemble the wild type riboswitch (Figure 16c).

It was found that crystallization was unable to occur with the other constructs as well as the control. Perhaps a Watson-Crick base pair between U96 and A33 was more favorable than formation of a base stack between G96 with A33. Trying to position G96 over the top face of A33 might have caused too much strain in the J5/PK2 region, thus destabilizing the pseudoknot.

Despite proposed base pairing between U96 and A33 in the 3’ UC overhang construct, it was thought that G32 would be too far away from C97 to form any favorable base stacking interactions. Because of this, the additional residue could have been mobile out in the solution, creating extra steric bulk that would have prevented favorable intermolecular lattice contacts. From here, it was validated that the 3’U overhang was crucial for stabilizing the pseudoknot formed.
Lattice Contacts

Nucleobases that are not base-paired or significantly interacting with other nucleobases can “flip-out” from the inside of the RNA structure and interact with the outside environment. These bulged nucleotides have the possibility to interact with exposed nucleotides of other neighboring RNA molecules. These points of interaction, known as lattice contacts, are important in overall stability of crystal structure. Because of this, nucleotide changes were made that were thought to help improve base stacking between bulged nucleotides of neighboring molecules in hopes to form a more well ordered crystal lattice. Because the 3’ U overhang proved to be critical in forming crystals, this mutation was implemented in all further mutant constructs.

Figure 16. a) Secondary structure of SAM-I/IV with 3’ UC overhang. b) Secondary structure of SAM-I/IV with 3’ G overhang. c) Secondary structure of control RNA without any 3’ overhang.
**A15G with 3’U Overhang**

After further review of the structure, it was found that there was a lattice contact between the A15 in the tetraloop of P2 with another A15 in a neighboring molecule (Figure 17a). Because of this, a mutation was made in this position to change the adenine to a guanine nucleotide (A15G) in order to create a GG base stack (Figure 17b). This was done in hopes to maximize favorable pi-stacking at this lattice contact. Crystals of prism shaped morphology (Figure 17c) were obtained and could be reproducibly grown in conditions of 0.05M Na-Cac pH 6, 0.03M MgOAc, 0.325M KCl, 8% MPD, 0.001M CoHex, while 20 μl of MPD diluted to 20% with well mother liquor was used as the cryoprotectant. While of diffraction quality, these crystals diffracted to a maximal resolution of 4.42 Å. It was thought that while the GG base stack might have provided some local stability, minute changes in orientation the molecules could have caused dis-favorable interactions between molecules at a global level, making the crystal lattice less ordered.
Figure 17. a) Molecular model of SAM-I/IV showcasing the intra-molecular base stacking of G15 with the G15 of a neighboring molecule (blue). b) Secondary structure of SAM-I/IV with A15 to G and 3’U overhang boxed in blue. c) Crystals of SAM-I/IV with the A15 to G change.
**C38G and A59G with 3’ U Overhang**

C38 was changed to a G to help stabilize an intramolecular base stack with A39. A59 was changed to G to help promote GG stacking with G81 of a neighboring molecule. Crystals for C38G and A59G were not able to be obtained (Figure 18 a-d). Perhaps ordering G38 into the RNA in order to base stack with A39 could have caused strain within this region of the RNA, which in turn would have made the overall crystal lattice less ordered. As with the change from A15 to G, stability of this one lattice contact (G59 and G81) could have negatively affected the overall crystal lattice such that it became less ordered.
Figure 18. a) Secondary structure of SAM-I/IV with C38 to G and 3’U overhang boxed in blue. b) Molecular model of G38 (cyan) base stacking with A39 (orange). c) Secondary structure of SAM-I/IV A59 to G with 3’U overhang boxed in blue. d) Molecular model of G59 (yellow) base stacking with G82 of neighboring molecule (magenta).
C7A with 3’U Overhang

C7 was changed to an A to help stabilize intramolecular base stacking with A6 (Figure 19a, b). Crystals of prism shaped morphology (Figure 19c) could be reproducibly grown in conditions of 0.05M Na-Cac pH 6, 0.045M MgOAc, 0.350M KCl, 8% MPD, 0.001M CoHex, while 20 μl of MPD diluted to 20% with well mother liquor was used as the cryoprotectant. While of diffraction quality, these crystals were worse in resolution quality than that of the A15G construct, which was lower than the single 3’ uracil overhang. The best resolution obtained was 5.14 Å. Results can be seen in Figure 15. While C7 and A6 don’t seem to be directly interacting with SAM, these two nucleotides appear to help provide structure for the binding pocket. Aligning these two bulged nucleotides could have hindered SAM from fitting correctly into the binding pocket. This could distort the overall structure of the molecule, and therefore, create a less ordered crystal lattice.
Figure 19. Crystal of C7A.  a) Secondary Structure of SAM-I/IV with C7 to A and 3’ U overhang boxed in blue.  b) Molecular model of SAM-I/IV showing G7 (blue) base stacking with A6 (magenta).  c) Crystal of SAM-I/IV with C7 to A and 3’U overhang.
GU was inserted back into the J5/PK2 region (Figure 20a). By inserting the GU nucleotides back in, it was thought that the addition of these two nucleotides would alleviate some stress in forming the pseudoknot by creating a Watson-Crick base pair between U92 in P5 with A39 in P3, as well as G93 in P5 with C38 in P3 (Figure 20c). It was hoped that this added stability in the pseudoknot would help molecular packing in the crystal lattice. Sharp, well-defined, diamond and prism crystal morphologies were obtained, making this construct the most diverse in observed crystal morphology (Figure 20b). Also, this construct appeared to have the best physical appearance in terms of edge definition and shape. These crystals were obtained in 0.05M Na-Cac pH6, 0.03M Spermine, 0.01M MgCl₂, 0.001M CuSO₄, 14% isopropanol, while 20 μl of MPD diluted to 20% with well mother liquor was used as the cryoprotectant. Despite the great physical appearance of the crystals, they diffracted quite poorly. The best resolution obtained was 7.43 Å. While the pseudoknot may have been stabilized, the alteration could have induced a significant structural change that would have poorly altered the overall crystal lattice.
Figure 20. a) Secondary structure of SAM-I/IV with GU insertion (magenta insert), A15 to G, and 3’U overhang. b) Crystals of SAM-I/IV with GU insertion, A15 to G and 3’ U overhang showcasing the variety of morphologies. c) Molecular model of SAM-I/IV with GU insertion, A15 to G and 3’ U overhang showcasing the pseudoknot. 3’ end is shown in green, G91 and U92 are shown in yellow, base pairs formed between the J5/PK2 (blue) and P3 (magenta) are shown.
Molecular Modeling

The deletion of G91 and U92 in the J5/PK2 region along with the 3’U overhang proved to be critical in obtaining diffraction quality crystals. This mutation, however, likely yields a different base pairing pattern than the pseudoknot found in the crystal construct. Because of this, a structure of the RNA including G91 and U92 back in to the J5/PK2 region needed to be modeled. It has been observed that loop-loop interactions, termed “kissing loops”, are structurally similar to loop-strand interactions, which are characteristic of pseudoknots. It was deemed acceptable to model the pseudoknot in the riboswitch to this homologous kissing loop structure.

Typically, these loop-loop and loop-strand interactions consist of six base pairs. The model constructed of SAM-I/IV only had four phylogenic base pairs along with the mutagenic U96 base pair to give a total of five base pairs in the pseudoknot region. L2 and L3 kissing loop of the *Thermatoga maritima* lysine riboswitch was deemed to be a suitable template to try and correct the un-natural pseudoknot. First, the loop extending from P3 of SAM-I/IV along with the last base pairs of P3 were completely removed (G32-C40). P2 of the lysine riboswitch was aligned on P3 of SAM-I/IV in order to superimpose L2 and the corresponding base pairs in L3 onto the PK2 region (Figure 21). The linker arm extending from P5 was then connected back to reform the J5/PK2 region. It was observed that the new pseudoknot model was well accommodated without clashing with any other parts of the RNA. This confirmed that the new pseudoknot was an excellent replacement for the old one. Finally, the base pair identities in the new pseudoknot were mutated back to the original pseudoknot sequence of SAM-I/IV.
The resulting model differs modestly from the original structure, with the primary difference being readjustment of the direction of the PK2 helix such that it is more contiguous with P3, notably the Watson-Crick base pair between U92 in P5 with A39 in P3, as well as G93 in P5 with C38 in P3. Thus, it is likely that the crystal structure is representative of the SAM-I/IV family (Figure 22).
Figure 22. a) Molecular model of the original pseudoknot which only consists of five base pairs. b) Molecular model of the Wild Type pseudoknot which consists of six base pairs. Molecular model of the SAM-I/IV riboswitch. 3' end is shown in green, 5' end in red, G91 is shown in yellow, base pairs formed between the J5/PK2 (blue) and P3 (magenta) are shown to portray the modeled pseudoknot. SAM ligand in the core is shown in orange.
Relationship between Sub-domains in the SAM-I Family

The model generated shows clear evidence that the core of SAM-I/IV is closely related to that of SAM-I. Since SAM-I, SAM-IV, and SAM-I/IV support different peripheries despite sharing a very similar core, an investigation was conducted to determine whether these sub-domains were mutually exclusive. Constructs from the RNA sequence used to solve the three dimensional structure of SAM-I/IV with the 3’ U overhang and the GU insertion into the J5/PK-2 region, as well as two new constructs consisting of both pseudoknots found in SAM-I and SAM-IV were made. The wild type sequence was also assessed and used for comparison.

Isothermal Titration Calorimetry

Isothermal Titration Calorimetry was used in order to assess the effect the mutations made in the pseudoknot had on binding affinity for SAM. These sets of experiments also gave insight into the significance of the two different pseudoknots formed in SAM-I and SAM-I/IV to determine whether these sub-domains were mutually exclusive.

Constructs made were Crystal Construct (the RNA sequence with 3’U overhang used to solve the preliminary three-dimensional structure), GU Insertion (the construct with the crystal structure sequence using the 3’ U overhang along with nucleotides G91 and U92 added back into the J5/PK2 region to model the Wild Type Pseudoknot), Wild Type (the wild type RNA sequence of env87), Hybrid Core I (the RNA sequence that contains the core binding pocket of SAM-I but also includes the pseudoknot of SAM-I and the pseudoknot of SAM-I/IV), and
Hybrid Core IV (the RNA sequence that contains the core binding pocket of SAM-I/IV but also includes the pseudoknot of SAM-I and the pseudoknot of SAM-I/IV) (Figure 23a and b).

**Figure 23.** a) Secondary Structures of the Wild Type, Crystal Construct, and the sequence used to model the Wild Type pseudoknot are shown. b) Secondary Structures of Hybrid Core I and Hybrid Core IV are shown.
All ITC experiments of each construct were conducted in two sets. The first set performed used RNA in a buffer solution containing 1 mM MgCl$_2$ to mimic intra-cellular conditions, while the other set used RNA in a buffer solution containing 10 mM MgCl$_2$ as way to assess maximal stability of the three dimensional RNA structure (See ITC sample preparation in Material and Methods Section). Results can be seen in (Table 3).

**Table 3.**

<table>
<thead>
<tr>
<th>Sequence</th>
<th>$K_d$ (μM)</th>
<th>n</th>
</tr>
</thead>
<tbody>
<tr>
<td>Crystal Construct 10</td>
<td>7.3 ± 0.4</td>
<td>1.04 ± 0.02</td>
</tr>
<tr>
<td>Crystal Construct 1</td>
<td>n.d.</td>
<td>n.d.</td>
</tr>
<tr>
<td>GU Insertion 10</td>
<td>0.15 ± 0.02</td>
<td>0.39 ± 0.02</td>
</tr>
<tr>
<td>GU Insertion 1</td>
<td>7.2 ± 2.6</td>
<td>0.58 ± 0.31</td>
</tr>
<tr>
<td>Wild Type 10</td>
<td>0.091 ± 0.021</td>
<td>1.1 ± 0.1</td>
</tr>
<tr>
<td>Wild Type 1</td>
<td>2.5 ± 0.5</td>
<td>0.89 ± 0.04</td>
</tr>
<tr>
<td>Hybrid Core VI 10</td>
<td>0.27 ± 0.03</td>
<td>0.81 ± 0.02</td>
</tr>
<tr>
<td>Hybrid Core VI 1</td>
<td>5.0 ± 3.0</td>
<td>0.1*</td>
</tr>
<tr>
<td>Hybrid Core I 10</td>
<td>0.081 ± 0.011</td>
<td>0.84 ± 0.03</td>
</tr>
<tr>
<td>Hybrid Core I 1</td>
<td>2.8 ± 0.2</td>
<td>0.1*</td>
</tr>
</tbody>
</table>

Data collected from ITC experiments from the different RNA constructs tested. $K_d$ is the dissociation constant and n is the stoichiometric ratio of ligand binding to RNA. Each experiment was done in triplicate and only the average values are reported. The standard deviation is reported next to each value if applicable. The numbers “10” and “1” next to each sequence denotes the concentration of MgCl$_2$ in buffer measured in mM.

*A very low n value was originally obtained for these experiments. An estimation was made that 10% of the RNA population in the sample cell was actually binding ligand. Because of this, n values were fixed to 0.1 to provide a more accurate measurement of $K_d$. *

The global trend showcases each construct reporting a lower $K_d$ at 10 mM MgCl$_2$ than at 1 mM MgCl$_2$. This is suggestive that the higher concentration of divalent ions in solution helps
increase the affinity of SAM binding. Because overall structure and stability of the riboswitch influences the structure of the binding pocket (which in effect is related to riboswitch binding affinity for ligand), then perhaps the higher concentration of divalent ions helps the riboswitch fold into a proper three-dimensional structure necessary for binding SAM.

The Crystal Construct is shown to have the lowest binding affinity for ligand at 10 mM MgCl$_2$. This could imply that the single 3’ U overhang does not provide enough stability for the pseudoknot which would cause the overall structure of the RNA as well as the structure of the binding pocket to be less ordered. At 1 mM MgCl$_2$ there was no detectable ligand binding, again reinforcing the notion that the 3’U overhang alone does not significantly aid in structural rigidity of the pseudoknot.

The construct made with the GU insertion had fairly similar binding affinities with the Wild Type sequence both at 10 mM MgCl$_2$ and 1 mM MgCl$_2$. This would be expected considering that the PK-2 region is nearly identical to that of the Wild Type sequence. However, the stoichiometric binding ratios are lower than that of the Wild Type. Even though the stoichiometric binding ratios of the GU insertion construct at both 10 mM MgCl$_2$ and 1 mM MgCl$_2$ have less than a three-fold difference between the Wild Type ratios at the same MgCl$_2$ concentrations, perhaps the flexibility of the un-paired 3’ U overhang could be clashing with other regions of the RNA, which could hinder its ability to sufficiently form a binding pocket. Further studies with this sequence that omitted the 3’ U overhang would need to be conducted in order to investigate this claim.
While it was expected that the Wild Type structure would have high binding affinity for SAM at 10 mM MgCl$_2$, it was interesting to see that Hybrid Core I and Hybrid Core VI also had very similar $K_d$ and stoichiometric binding ratio values at this divalent ion concentration, suggesting that both were able to bind ligand with a similar affinity as the Wild Type sequence. This was an interesting discovery in itself. Under these higher divalent ion concentrations, it was clear that the RNA was able to fold into a structure that allowed it to effectively bind ligand. This meant that these conditions allowed the core to support both sub-domains that were able to create a binding pocket with adequate recognition for SAM. Because a crystal structure of this mutant is not solved, the exact cause of how this folding event occurs is unknown. Further studies will need to be conducted in order to gain understanding of this phenomenon.

What is also interesting is that it was predicted that perhaps Hybrid Core I and Hybrid Core VI would have similar binding affinities for ligand because it was thought that the two nucleotide mutations (A7 and C67 in Hybrid Core I; C7 and U67 in Hybrid Core VI) in the binding pocket would not significantly influence binding affinity because these nucleotides do not directly contact SAM. Under conditions of higher ion concentrations, Hybrid Core I has close to a three-fold higher binding affinity for ligand than Hybrid Core VI. While not an alarming difference in affinity, perhaps the core of Hybrid Core I could be better suited for sustaining both sub-domains. C67 in Hybrid Core I forms a Watson-Crick base pair with G39, while Hybrid Core VI forms a base pair mismatch between U67 and G39. This Watson-Crick base pair in Hybrid Core I could provide the extra stability needed to organize the binding pocket to increase affinity for SAM over the UG mismatch. Further studies would need to be conducted in order to determine the true cause of the differences in affinity.
Hybrid Core I and Hybrid Core VI also show to have significant binding affinities for ligand. However, the original stoichiometric ratio for these binding events was so low that a parameter had to be set. An estimation put in place was that perhaps only a very small portion of RNA, such as 10%, was effectively binding ligand. With this estimation, the n value was set to 0.1 in order to try and determine a more reasonable $K_d$. While it was shown at 10 mM MgCl$_2$ that nearly all of the RNA was able to fold into the proper orientation to effectively bind SAM, the low stoichiometric binding ratio suggests that the majority of both Hybrid Core I and Hybrid Core VI constructs could not effectively form an adequate binding pocket for ligand. This could imply that while theoretically these constructs could bind to SAM given higher divalent ion concentrations that would aid with proper folding of the RNA, formation of an adequate binding pocket is certainly not feasible under cellular conditions. This could help suggest that under cellular conditions, these two sub-domains are in fact mutually exclusive for the SAM-I/IV family.

**Conclusions**

Condition space screening strategies proved successful to give crystal growth conditions that could reproducibly grow crystals of diffraction quality. These strategies can aid researchers who investigate crystal growth and how crystal quality can be improved by optimizing mother-liquor conditions instead of point-mutations within the RNA. Further research in this area could lead to a standardized protocol for mother liquors that could be implemented to all crystallographic studies that would allow for exact mother liquor conditions to give quality crystals to be solved for.
The mutations made for each construct also provided insight into how the stabilities of certain regions of the RNA affected crystal quality. The 3’ U overhang proved to be critical in crystal formation, suggesting the significant impact stabilization of the pseudoknot had on the rest of the RNA molecule as well as its influence on stabilizing the rest of the crystal lattice. The lower resolutions obtained with constructs where mutations were made within the sequence of the RNA suggested that stabilizing effects on one lattice contact can negatively affect the stability of other lattice contacts, which can be primarily seen with constructs A15G and A59G. Low X-Ray diffraction resolutions with constructs C38G and C7A also suggested that base stacking adjacent bulged nucleotides could strain certain regions of the RNA which could translate to less favorable intermolecular lattice contacts. Understanding how local changes in stability can influence overall molecular stability as well as lattice contact stability can aid in optimizing crystal quality as well as manipulating ligand binding affinity. Because riboswitches are becoming increasingly popular for drug targets, being able to “fine tune” the RNA’s affinity for small molecules will be advantageous in research for novel drug therapies.

Molecular modeling also proved to be a useful tool to provide details about the crystal structure from a lower resolution diffraction pattern. It was clear that addition of G91 and U92 gave the pseudoknot a substantially different structure compared with the pseudoknot without these additions. These structural differences proved to be critical in assessing the binding affinity for SAM.

The ITC studies have further underscored the effect of pseudoknot stability on ligand binding affinity. Low binding affinity for SAM with just a 3’U overhang in the pseudoknot
region suggests that a single base pair does not adequately provide the structural support needed to form an ordered binding pocket at cellular ion concentrations. Even though the 3’ U overhang was able to provide enough stability for the three dimensional structure to be solved, the high salt concentrations in the mother liquor could have significantly aided in stabilizing the RNA into the correct conformation. This could help show that while the pseudoknot does not directly interact with ligand binding, its stability helps provide adequate structure to create the binding pocket for SAM.

The results from the ITC experiments also showcase how Hybrid Core-I and Hybrid Core-VI can theoretically create three-dimensional architectures that allow for an ordered binding pocket to have high affinity for SAM given adequate stabilizing conditions. However, these two constructs did not adequately bind ligand under conditions of low MgCl₂, which suggests the intracellular environment could not provide sufficient aid to allow an RNA that supported these two peripheries to fold into the correct three dimensional arrangement necessary to bind SAM. This is highly suggestive that these two sub-domains are mutually exclusive under cellular conditions in the SAM-I/IV family.
### Appendix A

<table>
<thead>
<tr>
<th><strong>Oligo Name</strong></th>
<th><strong>Sequence</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>5’ Gen</td>
<td>GAAATTAATACGACTCACTATAGG</td>
</tr>
</tbody>
</table>
| 3’ U overhang  | A) GAAATTAATACGACTCACTATAAGATCACGAGGGAGACCCCGGCAAC  
                 B) GAGGGGGAGACCCCGGCAACCTGGGACGGACACCCAAGGTGCTCACACCAGGAGACGGTG  
                 C) TGCTCACACCCGGAGACGGTGGAATCCGGCCCGAGAGGGCAACGAAGTCCG  
                 D) mAmCGGACACTTCCGGTTCCTCT |
| 3’UC overhang  | A) GAAATTAATACGACTCACTATAGGATCACGAGGGGGAGACCCCGGCAAC  
                 B) GAGGGGGAGACCCCGGCAACCTGGGACGGACACCCAAGGTGCTCACACCAGGAGACGGTG  
                 C) TGCTCACACCCGGAGACGGTGGAATCCGGCCCGAGAGGGCAACGAAGTCCG  
                 D) mGmCGGACACTTCCGGTTCCTCT |
| 3’ G overhang  | A) GAAATTAATACGACTCACTATAGGATCACGAGGGGGAGACCCCGGCAAC  
                 B) GAGGGGGAGACCCCGGCAACCTGGGACGGACACCCAAGGTGCTCACACCAGGAGACGGTG  
                 C) TGCTCACACCCGGAGACGGTGGAATCCGGCCCGAGAGGGCAACGAAGTCCG  
                 D) mCmCGGACACTTCCGGTTCCTCT |
| A15G with 3’U overhang | A) GAAATTAATACGACTCACTATAGGATCACGAGGGGGAGACCCCGGCAAC  
                           B) GAGGGGGAGACCCCGGCAACCTGGGACGGACACCCAAGGTGCTCACACCAGGAGACGGTG  
                           C) TGCTCACACCCGGAGACGGTGGAATCCGGCCCGAGAGGGCAACGAAGTCCG  
                           D) mAmCGGACACTTCCGGTTCCTCT |
| C38G with 3’U overhang | A) GAAATTAATACGACTCACTATAGGATCACGAGGGGGAGACCCCGGCAAC  
                           B) GAGGGGGAGACCCCGGCAACCTGGGACGGACACCCAAGGTGCTCACACCAGGAGACGGTG  
                           C) TGCTCACACCCGGAGACGGTGGAATCCGGCCCGAGAGGGCAACGAAGTCCG  
                           D) mAmCGGACACTTCCGGTTCCTCT |
| A59G with 3’U overhang | A) GAAATTAATACGACTCACTATAGGATCACGAGGGGGAGACCCCGGCAAC  
                           B) GAGGGGGAGACCCCGGCAACCTGGGACGGACACCCAAGGTGCTCACACCAGGAGACGGTG  
                           C) TGCTCACACCCGGAGACGGTGGAATCCGGCCCGAGAGGGCAACGAAGTCCG  
                           D) mAmCGGACACTTCCGGTTCCTCT |
| C7A with 3’U overhang | A) GAAATTAATACGACTCACTATAGGATCACGAGGGGGAGACCCCGGCAAC  
                           B) GAGGGGGAGACCCCGGCAACCTGGGACGGACACCCAAGGTGCTCACACCAGGAGACGGTG  
                           C) TGCTCACACCCGGAGACGGTGGAATCCGGCCCGAGAGGGCAACGAAGTCCG |
<table>
<thead>
<tr>
<th>Add-91-GU with 3’ U overhang</th>
<th>D) mAmCGGACCTTCGGTGCCCTCT</th>
</tr>
</thead>
<tbody>
<tr>
<td>A) GAAATTAATACGACTCTATAGGATCACTATAGGACGAGGGAGACCCCGGCAAC</td>
<td></td>
</tr>
<tr>
<td>B) GAGGGGAGACCCCGGCAACCTGGGACGGACACCCAAGGTGCTCACACCGGAGACGGTG</td>
<td></td>
</tr>
<tr>
<td>C) TGCTCACACCGGAGACGGTGAGATCCGGAGGGCAACGAAGTGTCCG</td>
<td></td>
</tr>
<tr>
<td>D) mAmCGGACCTTCGGTGCCCTCT</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Add-91-GU with A15G and 3’ U overhang</th>
<th>A) GAAATTAATACGACTCTATAGGATCAAGGGGAGACCCCGGCAAC</th>
</tr>
</thead>
<tbody>
<tr>
<td>B) GAGGGGAGACCCCGGCAACCTGGGACGGACACCCAAGGTGCTCACACCGGAGACGGTG</td>
<td></td>
</tr>
<tr>
<td>C) TGCTCACACCGGAGACGGTGAGATCCGGAGGGCAACGAAGTGTCCG</td>
<td></td>
</tr>
<tr>
<td>D) mAmCGGACCTTCGGTGCCCTCT</td>
<td></td>
</tr>
</tbody>
</table>

References


