Selection and Structural Characterization of Novel Riboswitches From an Existing Riboswitch Scaffold

By: Makenna Morck
Department of Chemistry and Biochemistry
University of Colorado, Boulder

Defended April 3, 2014

Thesis Advisor:
Dr. Robert T. Batey, Department of Chemistry and Biochemistry

Thesis Committee:
Dr. Robert T. Batey, Department of Chemistry and Biochemistry
Dr. Joseph Falke, Department of Chemistry and Biochemistry
Dr. Nancy Guild, Department of Molecular, Cellular, and Developmental Biology

Table of Contents

Abstract

Riboswitches are gene regulatory elements found in the 5' leader sequence of many bacterial genes. In a riboswitch, RNA directly binds ligand to affect transcription or translation of a downstream gene. Such a system could potentially be exploited in synthetic systems. However, one of the major hurdles for creating synthetic riboswitch regulatory systems (and other RNA devices) is creating RNA aptamers that are not only selected for binding, but also have potential for regulatory activity. In the present study, this problem was addressed, utilizing the tertiary architecture of a known purine-binding riboswitch and reselecting the binding pocket for 5-hydroxy-L-tryptophan (5HW) binding using *in vitro* selection. Three families arose in the selection. Two of the families had single nucleotide deletions that destroy the tertiary architecture of the purine riboswitch, but the third family maintained the primary sequence elements critical for forming the tertiary interactions between P2 and P3. The aptamer was crystallized and its structure was determined, revealing conservation of the loop-loop interaction from the purine riboswitch, while the ligand-binding pocket was restructured to allow for 5HW binding. Ongoing research aims to determine the structure of a different family of aptamers that arose in the selection. Although this family does not resemble the original purine riboswitch, it does contain a previously characterized tryptophan aptamer that arose from a completely randomized selection. The aptamer is a small hairpin, so it was appended to the P3 and P4 helices of the SAM-I riboswitch to encourage crystallization. Currently, this SAM-I/tryptophan fusion RNA has been crystallized to a resolution of ~5.0 Å.

1. Introduction

In order to efficiently function and reproduce, all organisms must be able to sense stimuli in their environment and react to these signals. Traditionally, it was thought that these reactions were mainly regulated by protein factors, which were believed to possess the structural and informational complexity required to bind specific metabolites or respond to other stimuli and appoint a suitable response. However, it has become apparent that RNA is also capable of regulating gene expression based on a number of environmental stimuli.

Charles Yanofsky's groundbreaking discovery of transcriptional attenuation was one of the first observed RNA gene regulation mechanisms (Figure 1). The mechanism was originally recognized in the *trp* operon, which
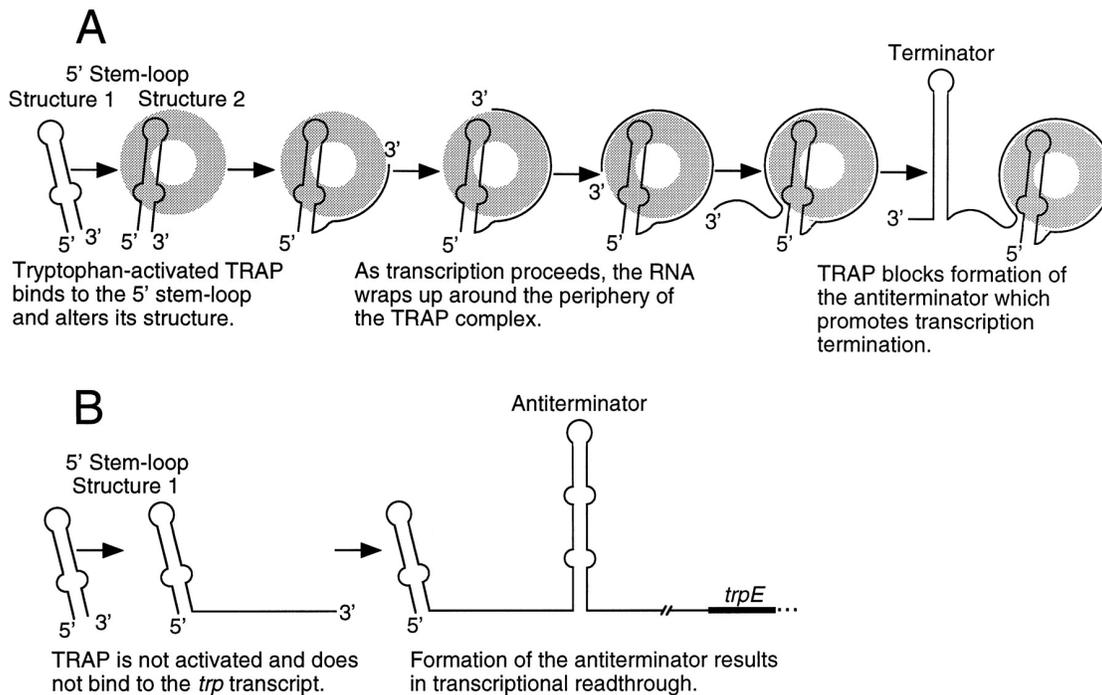
Figure 1: Transcriptional attenuation. A) In the presence of tryptophan, TRAP is activated and binds the nascent mRNA. The RNA wraps around the TRAP complex, preventing formation of the antiterminator stem and allowing for formation of the terminator stem. Transcription terminates. B) When tryptophan is not present, TRAP is not activated, and it does not bind the mRNA. The antiterminator is formed and the downstream gene is transcribed. Figure taken from Du et al., 2000.

controls tryptophan biosynthesis. A tryptophan RNA-binding attenuation protein (TRAP) was shown to control transcription based on intracellular tryptophan (Du et al., 2000). When TRAP is activated by high tryptophan concentrations, it binds the 5' leader sequence of a nascent mRNA and promotes formation of a terminator stem, terminating transcription (Du et al., 2000). In contrast, when tryptophan concentrations are low, the TRAP is inactive, and an antiterminator stem is formed in the mRNA, allowing gene read-through (Du et al., 2000).

Subsequently, it was discovered that RNA could directly bind a ligand and promote formation of the terminator or antiterminator stem without the need of a mediating attenuation protein. This new class of regulatory RNAs was collectively termed *riboswitches* (Winkler et al., 2002). Since their discovery, it has been demonstrated that over 4% of genes in *Bacillus subtilis* are controlled by riboswitches (Winkler and Breaker, 2005). Thus, for mechanisms that act after transcription has begun, RNA-dependent regulators are actually more frequent than protein-mediated regulation.

1.1 The Transcriptional Control Mechanism

Many riboswitch classes operate via the transcriptional control mechanism. These riboswitches can be divided into two distinct functional parts: the "aptamer," or sensing domain, and the "expression platform," or regulatory domain (Garst et al., 2009) As the riboswitch is transcribed, the aptamer domain can adopt one of two different conformations, dependent on the presence of its cognate metabolite. In an "OFF" switch, ligand binding stabilizes formation of the P1 helix, causing downstream formation of a terminator stem and aborting transcription. In the absence of ligand, an anti-terminator stem is formed, and the downstream gene is transcribed (Garst et al., 2009). "ON" switches have also been identified, where ligand binding allows read-through (Figure 2).
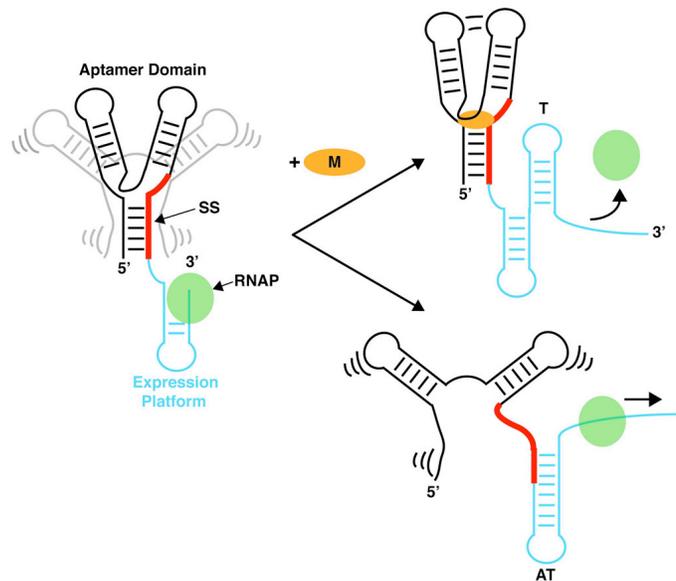
The structural characteristics of all major classes of riboswitches have been analyzed, elucidating the mechanisms RNA employs to bind a metabolite with high affinity and specificity. This interaction can be monitored *in vitro* via isothermal titration calorimetry (ITC), which results in an apparent K_D for ligand

binding. It can also be monitored *in vitro* using transcription assays that designate a T_{50}, which is the concentration of ligand required to elicit half-maximal response (Wickiser et al., 2005a). Additionally, *in vivo* assays have been developed which measure an EC_{50}, the concentration of ligand required to elicit half-maximal response *in vivo* (Ceres et al., 2013).

Aptamer domains can be categorized based on their secondary structures: type I, which encompasses riboswitches that bind their ligand in a multihelical junction, and type II, which are riboswitches that adopt an

Figure 2: Typical transcriptional riboswitch. In a typical transcriptional controlled riboswitch, in the presence of a metabolite (yellow), the switching sequence (SS) of the aptamer do main influences the formation of a Rho-independent terminator stem followed by a poly-U tract, which forces the RNA polymerase (green) to abort transcription of a downstream gene. In the absence of ligand, an anti-terminator stem is formed and read-through is permitted. Figure taken from Garst et al. 2009.

overall pseudoknot structure to bind the ligand (Serganov and Nudler, 2013). Type I riboswitches bind the ligand near the P1 stem and include including the thiamine pyrophospate (TPP) riboswitch (Edwards and Ferré-D'Amaré, 2006; Serganov et al., 2006; Thore et al., 2006), the lysine riboswitch (Garst et al., 2008; Serganov et al., 2008), and the purine riboswitch (Batey et al., 2004; Serganov et al., 2004). Type I also includes the tetrahydrofolate (THF) riboswitch, which binds a second ligand far from the P1 stem (Trausch et al., 2011). Representative type II riboswitches include the SAM II riboswitch (Gilbert et al., 2008) and the fluoride riboswitch (Ren et al., 2012).

The purine riboswitch is a well-understood example of riboswitch structure (Gilbert et al., 2006). It displays several RNA structural elements that appear

across riboswitch classes. One such element is coaxial stacking, apparent between the P1 and P3 helices, which creates a single pseudo-continuous helix (Garst et al., 2011). Additionally, the helices are arranged in a parallel conformation, which is apparent in most RNAs with three or more helices (Holbrook, 2008). The caps of the P2 and P3 helices, L2 and L3, are arranged in a kissing-loop motif (Gilbert et al., 2006).

1.2 Alternative Regulatory Mechanisms

In addition to the Rho-independent transcriptional control mechanism, riboswitches act in many alternative methods to elicit a regulatory response. Recently, Rho-dependent transcriptional control was observed in a Mg^{2+}-sensing riboswitch (Hollands et al., 2012). This mechanism had previously been suggested through genomic analyses that predicted the presence of several riboswitches but lacked the primary sequence required to function as Rho-independent transcriptional terminators (Irnov et al., 2010, Livny and Waldor, 2010, Rhagavan et al., 2011).

Additionally, many translationally controlled riboswitches have been identified (Lu et al., 2008, Rieder et al., 2007, Johnson et al., 2012). These riboswitches typically act through the sequestration of a ribosome binding sequence (RBS), preventing the ribosome to from initiating translation.

Other mechanisms have been demonstrated in a select few riboswitches. Among these, the thymine pyrophosphate (TPP) riboswitch is the only riboswitch that has been identified in both eukaryotes and bacteria. Because eukaryotes have very different requirements for transcriptional and translational gene control, the TPP riboswitch is involved in alternate gene splicing mechanisms (Cheah et al, 2007; Wachter et al., 2007). In green algae, the TPP riboswitch creates a prematurely terminated peptide (Croft et al., 2007). In filamentous fungi, when ligand is bound, alternate splicing causes the mRNA to produce an aberrant peptide, which is then degraded by the cell (Cheah et al., 2007). The TPP riboswitch is also found in some higher plants, where alternative splicing causes

unstable long 3' untranslated regions (Bocobza et al., 2007, Wachter et al., 2007).

1.3 Applications In Synthetic Biology

Due to their ability to directly sense small molecules and affect a regulatory response without the need for additional mediating proteins, riboswitches are a particularly attractive target for creating novel gene-control elements. Synthetic biology is a growing field due to an increasing demand for renewable chemical synthesis mechanisms (Keasling, 2012). Naturally occurring enzyme pathways have been stitched together to create novel synthetic systems (Keasling, 2012), but such synthetic pathways may require optimization of transcription levels of each gene. For example, in designing a synthetic system for the synthesis of artemesinin (an anti-malarial drug) a natural mevalonate pathway was utilized, but enzyme levels were unbalanced, causing a buildup of HMG-CoA (Keasling, 2012). To alleviate this problem, mRNA secondary structures were designed to differentially stabilize mRNAs in the pathway and balance enzyme levels (Keasling, 2012). However, the design of a synthetic riboswitch could be an even more elegant mechanism to control gene expression levels. With a synthetic riboswitch, buildup of any intermediate could be monitored by a riboswitch that senses that product, and the enzyme that synthesizes that product could be downregulated, or the next enzyme in the pathway could be upregulated.

To date, RNA engineering approaches have focused primarily on creating independent input or output domains. Output domains have been successfully created to induce fluorescent signal upon ligand binding using the malachite green aptamer (Stojanovic & Kolpashchikov, 2004). The fluorescent reporter domain of the RNA is fused to the ligand-sensing domain via a communication module composed of a short helix (Figure 3). Ligand binding increases the stability of the communication module, allowing the fluorescent reporter domain to fully form and bind the fluorescent ligand, increasing fluorescence (Stojanovic & Kolpashchikov, 2004). Additionally, synthetic gene-regulatory output domains

have been constructed by fusing aptamer domains to a non-native expression platform (Trausch et al., 2011, Ceres et al., 2013a). These chimeric riboswitches obey simple secondary-structural principles that can easily be optimized to create a switch with the desired regulatory response to a particular metabolite concentration (Ceres et al., 2013a).

Figure 3: Malachite green fluorescent reporter. The malachite green aptamer is fused to an ATP-binding aptamer via a communication module. Upon ATP binding, malachite green binding is promoted, and fluorescence increases. Figure adapted from (Stojanovic & Kolpashchikov, 2004).

However, these approaches have only proven broadly applicable for naturally occurring aptamer domains, which have been identified for a relatively small subset of cellular metabolites, let alone other molecules. RNA aptamers that bind any metabolite of interest can readily be selected using *in vitro* selection techniques (Ellington and Szostak, 1990). Some of these synthetic RNA aptamers have been fused to output domains to generate a response, but this approach has typically been slow, requiring tedious optimization (Ceres et al., 2013, Berens et al., 2001). The caveat in this approach may be that in an *in vitro* selection, the aptamer is selected exclusively for its ability to bind a small molecule, whereas in a natural riboswitch, upon ligand binding, the aptamer domain must communicate to the expression platform and regulate transcription.

The present study aimed to develop a novel riboswitch capable of binding a non-native ligand and regulating gene expression. To this end, the ligand-binding pocket of the purine riboswitch was reselected for binding to a novel ligand, while distal residues were left intact. The tertiary structure of the purine riboswitch has been selected by nature to fold robustly and on a transcriptionally relevant timescale. By leaving the overall structure of the purine riboswitch undisturbed, it was hypothesized that the RNA could maintain regulatory abilities,

even while binding a novel ligand. The purine riboswitch was reselected for 5-hydroxy-L-tryptophan (5HW) binding via *in vitro* selection. A resulting aptamer sequence was crystallized, revealing remarkable structural conservation to the original purine riboswitch, while selectively binding 5HW. Current studies aim to crystallize a different aptamer sequence that resulted from the selection in order to better understand how different aptamers can bind the same ligand with different affinities and specificities.

2. Selection of a Novel Aptamer from an Existing Riboswitch Aptamer Domain

The aptamer domain of a purine riboswitch was reselected for binding to 5-hydroxy-L-tryptophan (5HW), a serotonin precursor. 5HW was chosen not only for its potential applications in the serotonin biosynthetic pathway, but also for the number of potential ways an RNA could recognize and bind it. 5HW has four functional groups that can form hydrogen bonds: the exocyclic 5-hydroxyl group, the endocyclic amine, and the amino and carboxyl groups. Additionally, the tryptophan ring system provides a good surface for pi-stacking interactions, which are often utilized in nucleic acid ligand recognition.

2.1 Methods

2.1.1 Library Preparation

In order to alter ligand binding in the purine riboswitch, 23 residues located within 8 Å of the binding pocket were selected for randomization (Figure 4). Thus, the P1 stem and the loop-loop interaction between P2 and P3 was left intact, while residues in the three-way junction binding site were randomized. Sequences
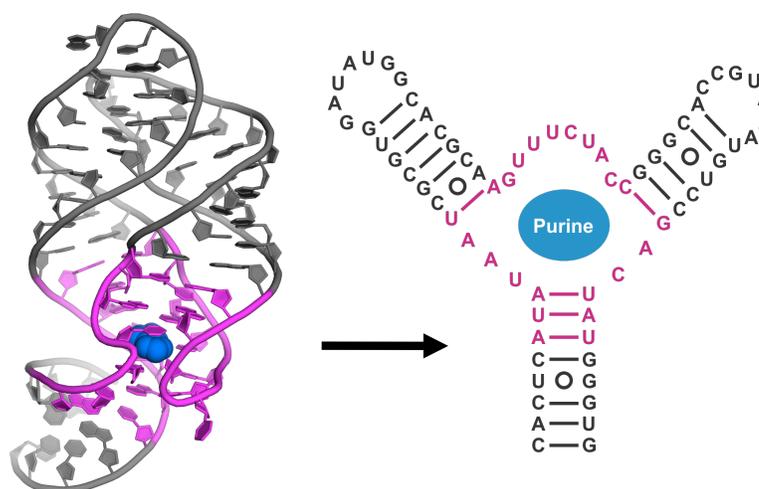
Figure 4: Randomized regions of the purine riboswitch. Nucleotides within 8 Å of the ligand-binding pocket (magenta) were randomized to create a starting library. Structure adapted from Gilbert et al., 2009 (PDB ID: 3FO6).

were constructed with the aptamer domain flanked by non-variant SHAPE cassettes as described in Wilkinson et al., 2006. These SHAPE cassettes allowed for PCR amplification of the randomized library after each round of selection. Oligomer sequences are listed in supplemental information. The resulting library had 4^{23} = 7 × 10^{13} unique sequences. During the selection, 200

pmol (1.2×10^{14} sequences) of RNA were present in the initial library. Thus, there was almost twofold redundancy of sequences in the initial library over unique sequences available in the selection. This ensured that every potential sequence was likely present in the initial pool, such that that no potential binders were omitted. This would not be possible in a selection with more randomized nucleotides. Furthermore, each sequence in the pool contained the information required to allow robust folding of the tertiary kissing loop from the purine riboswitch, creating a well-structured starting library.

The library was tested for primer extension efficiency as described in Piasecki et al., 2009, and it was determined that the library extended with 10% efficiency. Thus, in order to ensure PCR amplification of every unique sequence, an initial PCR[1] was performed with 10 µL of 37 µM randomized DNA template. The resulting DNA was used as the starting library for *in vitro* selection.

The DNA was transcribed *in vitro* with T7 polymerase and purified in a 6, 8, or 12% polyacrylamide denaturing gel. The RNA was then concentrated using Millipore 10,000 molecular weight cutoff (MWCO) concentrators and refolded by denaturing in 8 M urea and exchanging into a selection buffer containing 250 mM sodium chloride, 10 mM magnesium chloride, 50 mM potassium chloride, and 10 mM HEPES pH 7.0. In addition, a 125 µL transcription was performed with 2 µL radiolabeled α-ATP, creating a smaller labeled RNA library. The radiolabeled library provided a marker to indicate the approximate amount of RNA washed off of the column versus the amount of RNA eluted. A smaller radiolabeled marker library was similarly created after each subsequent round of selection.

2.1.2 *In vitro* Selection

The following selection protocol was performed in parallel with my graduate student mentor, Ely Porter. This was done in order to demonstrate that

---

[1] Unless otherwise noted, all basic RNA preparation and purification techniques used throughout the thesis research (including polymerase chain reaction techniques, *in vitro* transcription techniques, and denaturing polyacrylamide gel electrophoresis) were performed as described in Edwards et al., 2009.

two independent iterations of the same selection could result in the same RNA aptamer sequence.

To select for 5HW binding, a column containing sepharose beads linked to the carboxyl terminus of 5HW was synthesized. Before the first and second rounds of selection, a counter-selection was performed against acetylated sepharose. During column synthesis, any beads unmodified by 5HW were capped with an acetyl group. Counter selection removed any sequences that preferentially bound the sepharose bead, the linker between the sepharose and the amino acid, or the uncapped end of a linker with an acetyl group.

Five rounds of selection were performed, where 200 pmol of unlabeled RNA in 270 µL of selection buffer were added to 80 µL of radiolabeled RNA and loaded onto the equilibrated column. At the start of each round of selection, the RNA incubated at room temperature on the column for 45 minutes. RNA flow-through was collected by centrifugation. The RNA was then washed 3 – 10 times (3 in round 1, 6 in round 2, and 10 each in rounds 3 – 5) with 350 µL of selection buffer. Two hour elutions were performed, collecting RNA two or four times (two times in rounds 1 – 3, and four times in rounds 4 and 5) over the course of the selection.

An elution profile was created by blotting 10 µL of each flow-through, wash, and elutant in separate spots onto an electrostatic DEAE membrane. The amount of radiolabeled RNA in each spot was imaged with a Typhoon scanner and quantified using ImageJ, a free imaging program capable of integrating the density of each spot (Figure 5).
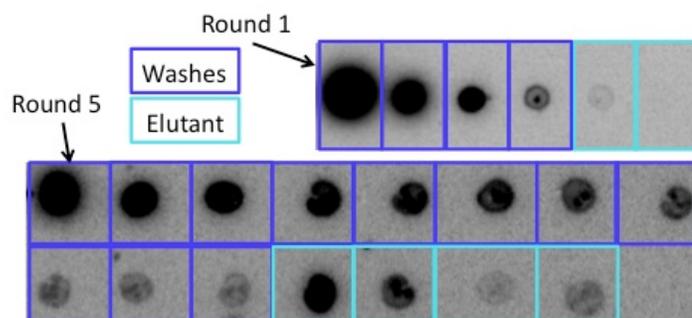

Figure 5: Quantification of selection rounds. Radiolabeled RNA is imaged and quantified after each round of selection. Total elutant (cyan) was summed and divided by total RNA.

Dividing total elutant RNA by total RNA flow through for each round resulted

in a measure of percent RNA elution. For example, in round 1, two elutions (E1 and E2) and three washes (W1, W2, and W3) were performed, in addition to the column flow-through (FT):

(E1 + E2) / (FT + W1 + W2 + W3 + E1 + E2) = Fraction elution

At the end of each round, the elutant RNA was pooled and reverse transcribed to create a DNA library for the next round of selection. SuperScript III reverse transcriptase (RT) was obtained from Invitrogen and was utilized as
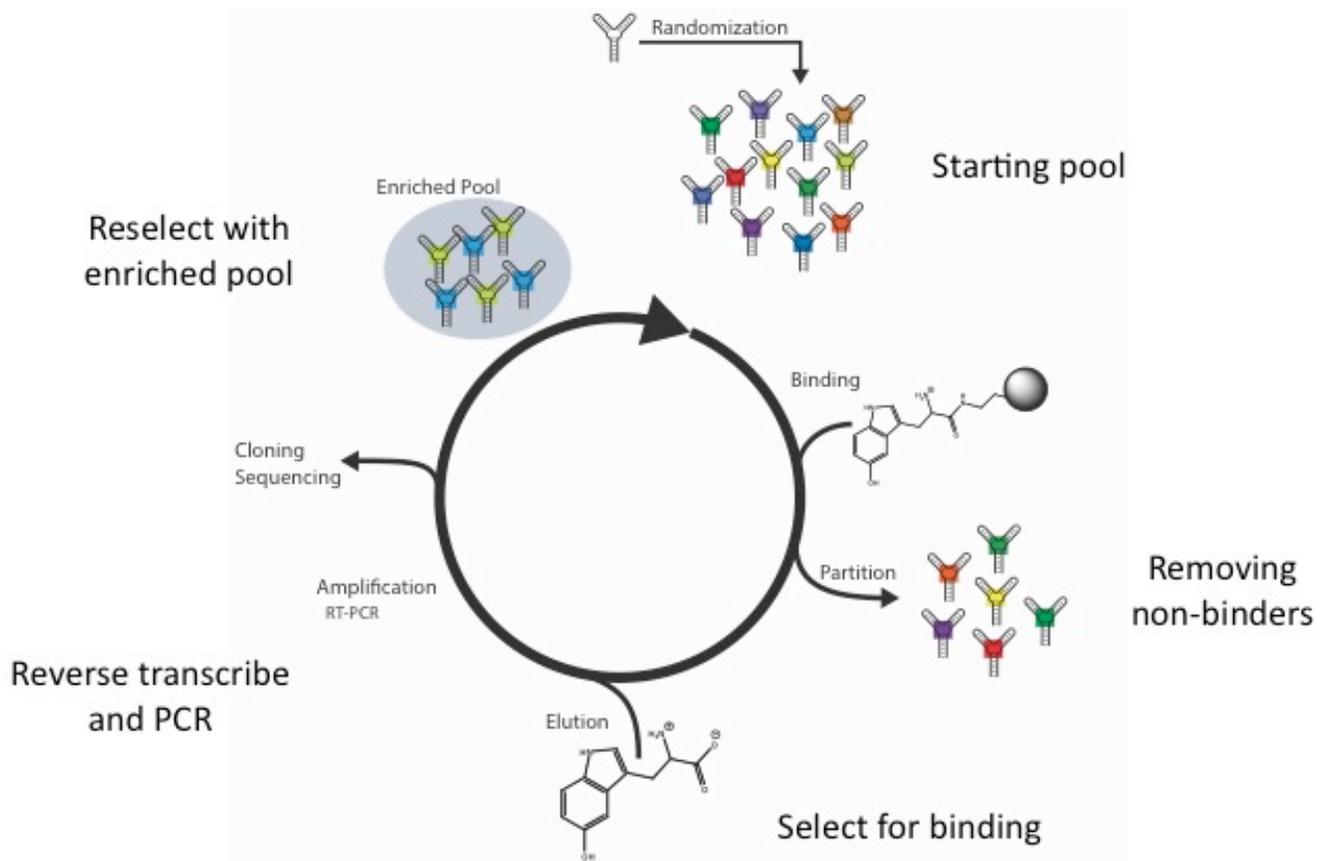

Figure 6: *In vitro* selection technique. 5 rounds of selection were performed. RNAs were selected for binding to 5HW linked sepharose and competitively eluted with free 5HW. Eluted RNA was then reverse transcribed and PCR amplified before transcription, after which they were subjected to another round of selection. Image courtesy Ely Porter.

described in protocols provided. The resultant ssDNA library was PCR amplified, transcribed, and purified as before and subjected to subsequent rounds of selection (Figure 6).
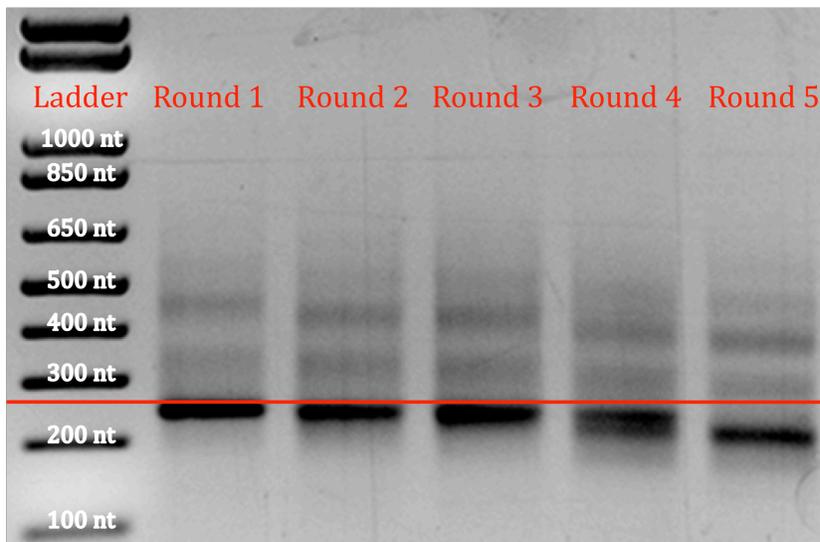
2.2 Results

        Significant elution was detected by round 3 of the selection, and by the end of round 5, almost 20% of the RNA could tightly bind 5HW (Figure 7).
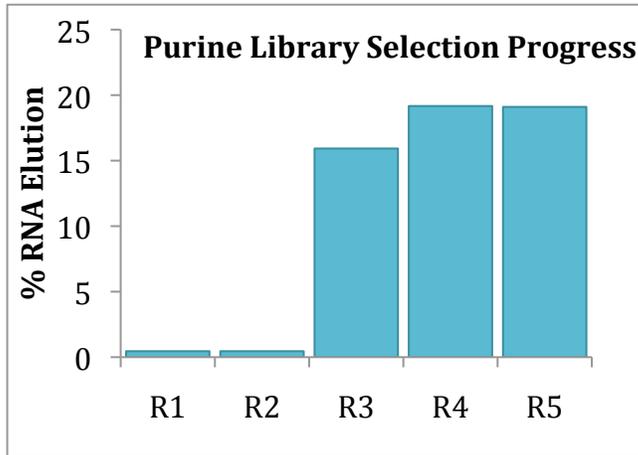
Figure 7: RNA elution after each round of selection. RNA elution was calculated as a percentage of total RNA. Significant elution was detected after the third round of selection, and rounds 4 and 5 resulted in almost 20% elution.

However, further analysis of the resultant DNA revealed problems with the final library. When DNA from each of the 5 rounds of selection was separated on a gel, it became apparent that there were deletions of about 20-50 nucleotides beginning in round 4 of the selection (Figure 8). The shortened species became dominant after round 5, suggesting that the deletions allowed the RNA to better survive the selection pressure.


Figure 8: DNA amplification of each selection round. Gel electrophoresis of DNA from each round in the selection demonstrates large deletions of 20-50 nucleotides between rounds 3 and 5 (visible in the main product band). Additionally, many longer products are present. Both of these errors can be attributed the reverse transcriptase, which is derived from a viral RT.

Reverse transcriptase is the most likely culprit for such deletions, given that SuperScript III is derived from a viral RT, M-MLV (Moloney Murine Leukemia Virus Reverse

Transcriptase). Viruses often favor error-prone polymerases in order to rapidly increase their genetic diversity. SuperScript III is not well equipped to proceed through structured RNAs with high fidelity. Because our selection began with a structured library, the RT likely introduced deletions. These errors propagated because the RT most efficiently replicated less-structured RNA variants.

Additionally, RT is known to strand-switch from the 5' end of an RNA to the 3' end of a new RNA without releasing the nascent DNA (Luo, 1990). This accounts for the longer species visible in Figure 8. A size selection was performed after each round of transcription, preventing the longer species from propagating through the selection. However, smaller changes in product length
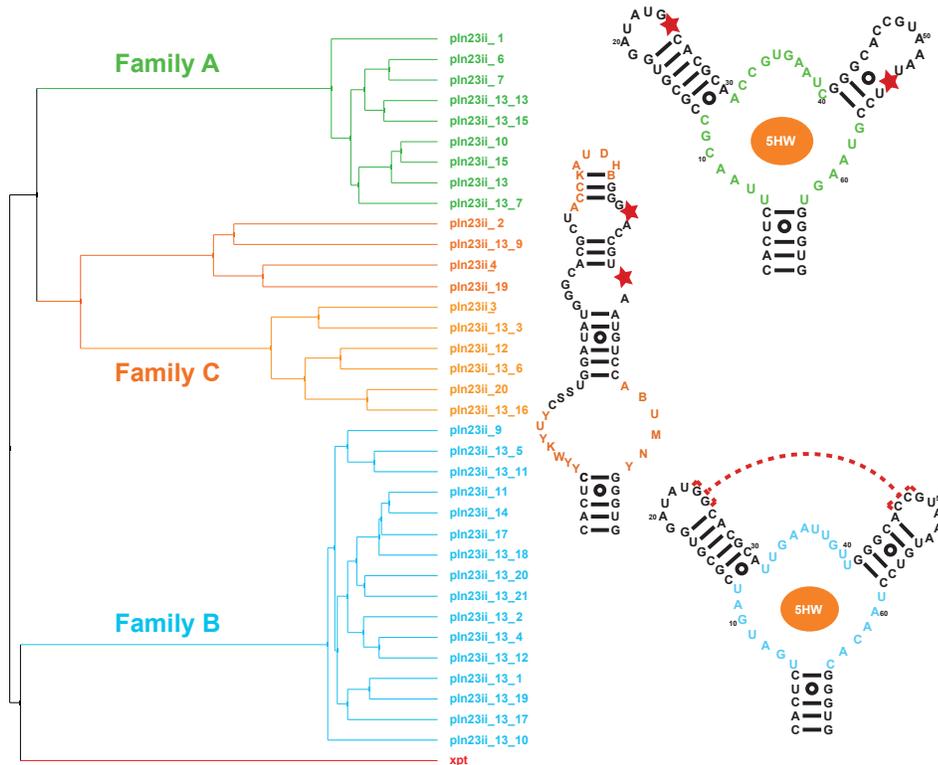
Figure 9: Phylogenetic tree of round 7 sequences. Ely's selection was analyzed after round 7; three main families were represented. Families A and C contained deletions destroying the tertiary architecture of the original purine aptamer sequence (labeled here "xpt"), but family B maintained the tertiary architecture of the purine riboswitch. Thus, family B was selected for further characterization. Secondary structures are predicted by minimum free energy modeling.

(such as the 20-50 nucleotide deletion) were not detected in the size selection step, and could thus proliferate. Further experiments by other Batey lab members will investigate the potential use of non-viral RT polymerases to decrease error in any future selections.

Despite the problems I faced in my selection, Ely was successfully able to complete his selection from the same starting library through 7 rounds, including counter-selections in rounds 6 and 7 against related ligand L-tryptophan. Phylogenetic analysis of round 7 sequences resulted in 3 aptamer families (Figure 9).

Families A and C both had two single-nucleotide deletions that destroyed the tertiary interactions of the original purine riboswitch. Family A deletions occurred in the capping loops of P2 and P3, destroying the tertiary loop-loop interaction. These single-nucleotide deletions may again be signatures of the reverse transcriptase. Alternately, family C deletions allowed the RNA to adopt a completely different secondary structure (as predicted by minimum free energy models). The RNA is predicted to form a single long hairpin with several loosely structured regions. Once more, the RT would likely favor this structure and propagate these deletions through the pool.

Family B aptamers, however, maintained the sequence arrangement of the original purine riboswitch in non-randomized sections while adopting an entirely unique sequence in randomized regions. Additionally, Ely was able to demonstrate via isothermal titration calorimetry (ITC) that the family B sequence bound 5HW with high affinity and specificity: The family B aptamer bound 5HW with a K_D of 3.9 ± 0.1 μM, while it bound L-tryptophan with a K_D of 280 ± 30 μM. Thus, the family B aptamers were selected for further characterization.

3. Crystallization and Structure Determination of Family B Aptamer

Since family B emerged as the best aptamer from the selection, its crystal structure was determined in order to demonstrate that it maintained the tertiary architecture of the purine riboswitch in non-randomized regions, and to understand its ability to recognize 5HW with a high affinity and specificity.

3.1 Crystallization
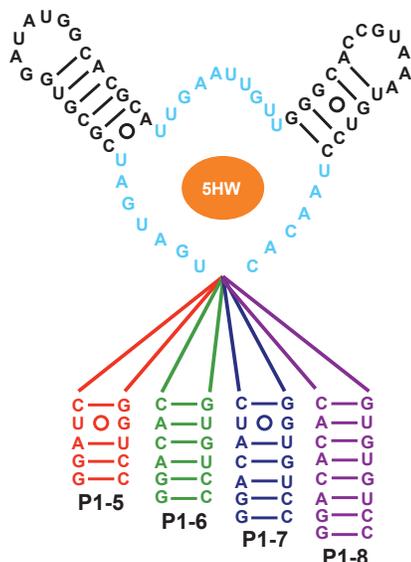
Figure 10: Family B crystallization constructs. Four sequences were designed and tested for crystallization. Each sequence was invariant in P2, P3, and the randomized regions, but had a different P1 stem length. P1-6 crystallized most readily.

To determine the structure of the family B aptamer, four different variants of family B were constructed and tested for their ability to crystallize. Each RNA was constant in both the randomized regions and the P2 and P3 helices, while the P1 stem length was altered to lengths 5, 6, 7, and 8 base pairs (Figure 10). Sequence of the P1 stem was chosen based on previous successes in the Batey lab in crystallizing the purine riboswitch (Batey et al., 2004). The sequences were then PCR amplified, *in vitro* transcribed, and purified via 8 or 12% denaturing polyacrylamide gel. The RNA was concentrated in Millipore 10,000 MWCO concentrators and exchanged into .5X Tris-EDTA buffer.

The standard hanging-drop method was used to crystallize the RNA. In hanging drop crystallography, a drop containing roughly equal parts RNA in buffer and a mother liquor solution is suspended over a solution of pure mother liquor in a sealed container. The mother liquor contains solvents in which RNA is generally insoluble and/or high salt concentrations. Vapor diffusion causes water to move out of the drop and mother liquor to move into the drop such that the

vapor pressure of the drop and the mother liquor below eventually equalize. This slowly decreases the solubility of the RNA, allowing crystallization to occur.

Crystallization of the four constructs was first tested using kits obtained from Hampton Research, including the Nucleic Acid Mini Screen, the Natrix 1 and 2 Screens, and the Crystal Screen. Crystal trays were set up as directed by protocols contained in each kit using RNA solutions 2 mM in 5HW and 400 µM in RNA in .5X Tris-EDTA buffer. The RNA with P1 stem length 6 was observed to generate robust crystals in Nucleic Acid Mini Screen condition 3 (Figure 11).
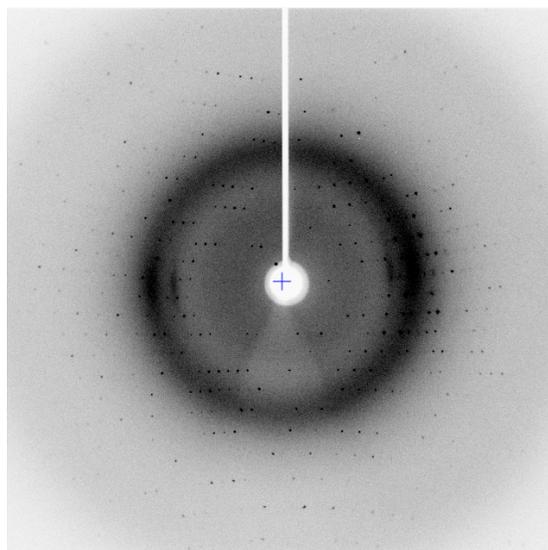
A   **B**
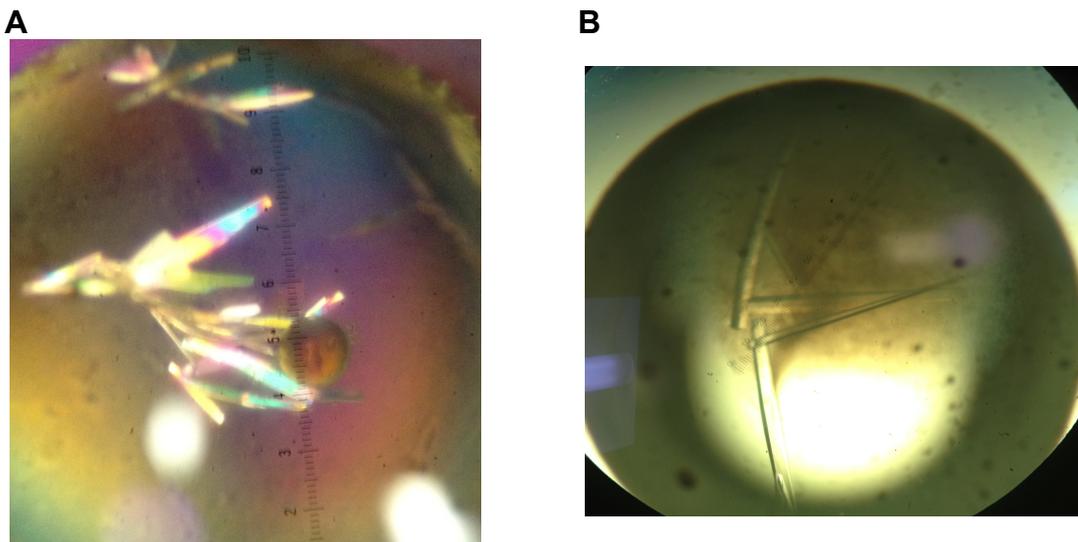
Figure 11: Nucleic Acid Mini Screen condition 3, P1-6 RNA. A) P1-6 readily crystallized in condition 3 from the Nucleic Acid Mini Screen. **B**) The mother liquor conditions were optimized to obtain diffraction-quality crystals.

Conditions were further optimized such that diffraction-quality crystals were reliably formed (Figure 11). The best crystals were produced in 8-14% MPD, 40 mM sodium cacodylate pH 5.5, 4 mM magnesium chloride, 12 mM sodium chloride, 80 mM potassium chloride, and either 4 – 9 mM cobalt hexammine or 1 – 11 mM iridium hexammine. Each drop contained 2.0 µL RNA solution and 3.5 µL mother liquor, and crystals were generally fully formed after 2 – 3 days at 22 °C. Additionally, a micro-seeding technique was employed to encourage nucleation: Drops with lower-quality crystals were harvested, and crystals were crushed under a microscope before transfer to a microcentrifuge tube. Cat whiskers were dipped into the solution and streaked across each drop

during assembly of a new tray, depositing very small amounts of crystalline RNA. This provides nucleation points on which new, high-quality crystals can assemble.

Crystals were cryoprotected by the addition 10 μL mother liquor to the crystal drop. The crystals were then mounted in a nylon loop and screened on the home source x-ray beam: a Rigaku MSC x-ray diffractometer with a rotating copper anode. Crystals diffracted to a resolution of 2.0 Å and fit into the C2 monoclinic space group in the presence of both iridium hexammine and cobalt hexammine (Figure 12). The structure was solved using the single isomorphous replacement with anomalous signal (SIRAS) method, where cobalt

Figure 12: Typical diffraction pattern of family B P1-6 crystals. The crystals diffract to a resolution of ~2.0 Å. The image shown corresponds to an RNA co-crystallized with iridium hexammine.

hexammine was replaced with iridium hexammine, which in turn has a small amount of anomalous scattering at the copper anode wavelength. R_{work} and R_{free} were determined to be 0.22 and 0.26, respectively, which fall under the acceptable limits.
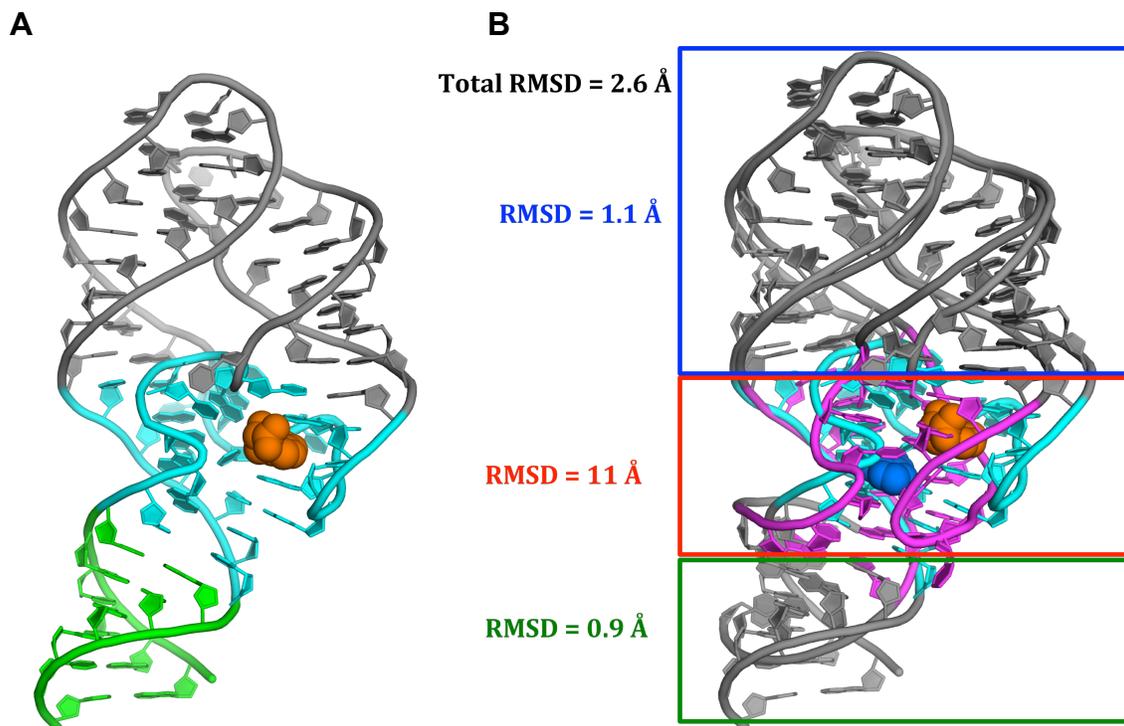
3.2 Structure

The structure of the family B aptamer displayed remarkable similarity to the original purine riboswitch (Figure 13). In particular, the loop-loop interaction between P2 and P3 is almost identical, with a root-mean-square difference (RMSD) of 1.1 Å as calculated by PyMOL, a free molecular modeling software. The P1 stems overlay with an RMSD of 0.9 Å, while the binding pockets overlay with a relatively large RMSD of 11 Å, indicating essentially no relationship. Overall, the molecules have an RMSD of 2.6 Å, reflecting the structural similarity

of the RNAs in non-randomized regions and the almost complete rearrangement of the binding pocket.

A   **B**

Figure 13: Crystal structure of the family B riboswitch. A) Family B structure, with randomized regions in cyan, 5HW in orange, and the P1 helix in green. **B**) Family B structure overlaid with the original purine structure. Root-mean square deviations for the P2-P3 interaction (blue), randomized region (red), and P1 helix (green) were calculated independently of the rest of the structure. Total RMSD (black) was calculated to be 2.6 Å.

The binding of 5HW is mediated by U13 in the junction between the P1 and P2 helices (J1/2), as well as residues U34, A36, and A37 in the junction between the P2 and P3 helices (J2/3) (Figure 14). The RNA recognizes all of the major functional groups of 5HW, which likely contributes to its high affinity and specificity. The amino group of 5HW forms hydrogen bonds with both the 2'-hydroxyl on the ribose ring of U13 and the endocyclic amine on A37. The 2-carbonyl group of U34 forms a hydrogen bond with the endocyclic amine on 5HW. The 5HW ring system forms pi stacking interactions with A36 and A37. Additionally, the exocyclic 5-hydroxyl group in 5HW makes a backbone contact to the phosphate group of U39. The RNA is selective for 5HW over tryptophan, and

this backbone contact is likely responsible. One of the oxygen atoms in the carboxyl group of 5HW forms a hydrogen bond with the exocyclic amino group
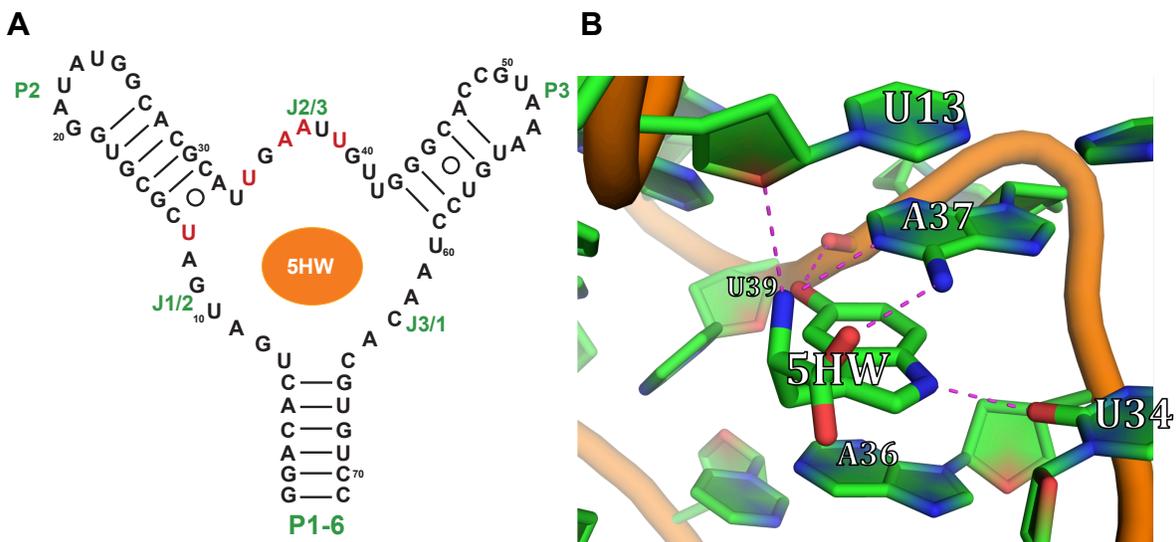
Figure 14: 5HW binding in the family B aptamer. A) Secondary structure of the family B riboswitch, with U13, U34, A37, A37, and U39 shown in red. These residues are primarily responsible for 5HW binding. **B)** Tertiary structure of the family B ligand-binding pocket. All of the major functional groups of the 5HW are recognized through hydrogen bonding, with the exception of the carboxyl terminus, where the linker stem would have been located. Additionally, 5HW forms pi-stacking interactions with both A36 and A37.
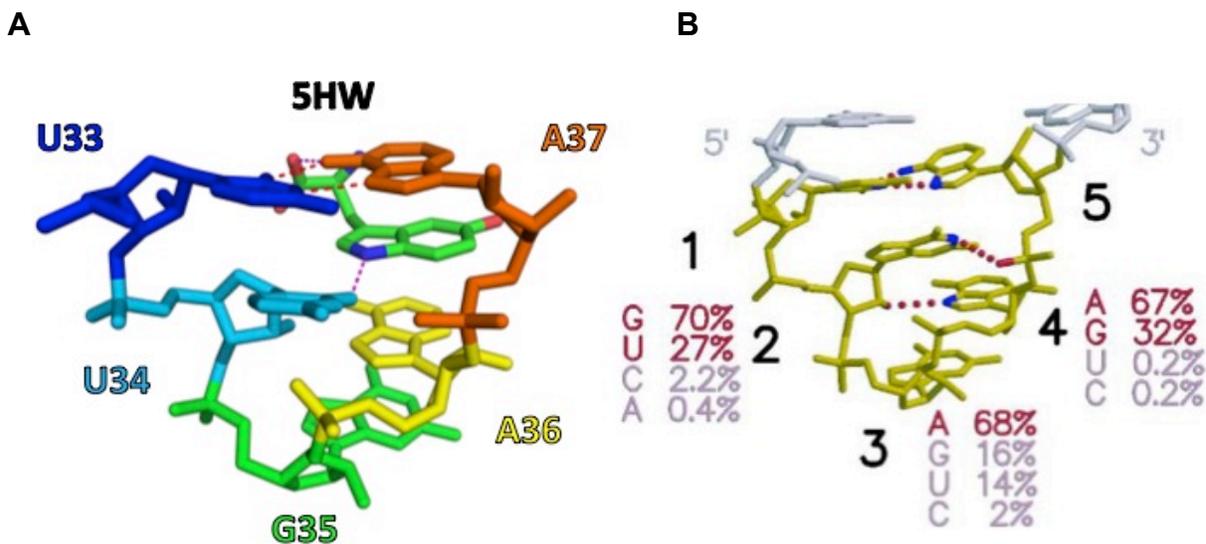

Figure 15: T-loop motif is exploited by 5HW. A) The T-loop motif in the 5HW-binding pocket of family B allows 5HW to pi-stack between A36 and A37. **B)** The T-loop motif is a common RNA structural element (Krasilnikov & Mondragon, 2003).

on A37, while the other oxygen atom is noticeably unrecognized. The unrecognized oxygen atom is where the linker stem to the sepharose bead would have been bound in the selection.

The main binding region from U33 to A37 forms the familiar T-loop motif (Krasilnikov & Mondragon, 2003) (Figure 15). The T-loop motif is a common RNA formation found in many RNA structures. It was first identified in tRNA (Krasilnikov & Mondragon, 2003). U33 and A37 are base-paired, while U34, G35, and A36 form a sharp turn. 5HW exploits base stacking between A36 and A37, binding between the two bases. Overall, the structure corroborates previous evidence that the aptamer can maintain the overall tertiary architecture of the purine riboswitch while selectively binding 5HW.

## 4. Current Research: Crystallization of the Family C Aptamer

Although family B was the only aptamer that maintained the tertiary architecture of the purine riboswitch, the family C aptamer was also of interest. Of particular note was that part of the predicted secondary structure matched the consensus sequence of a previously characterized tryptophan-binding aptamer that was derived via a completely randomized selection (Majerfeld et al., 2010) (Figure 16). Upon further characterization of family C, Ely was able to determine
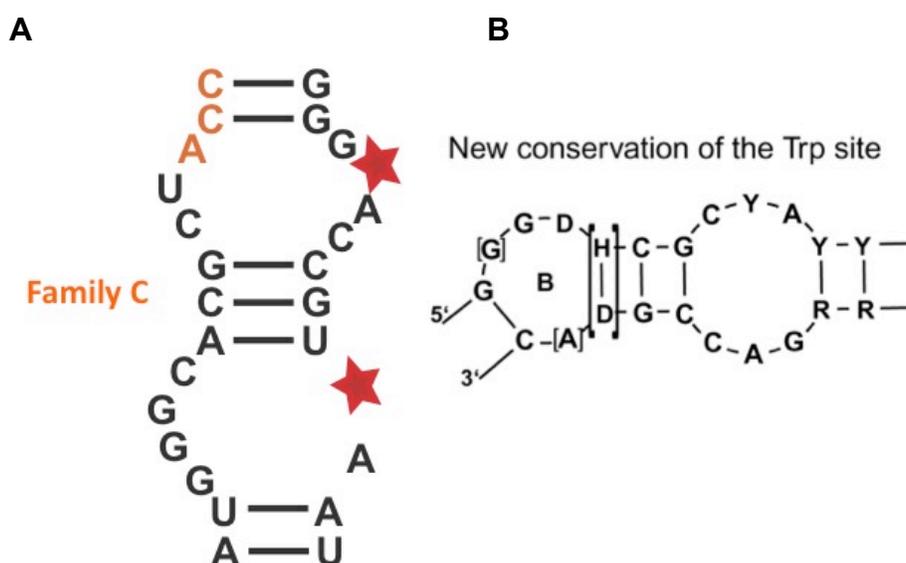
Figure 16: Family C is a previously characterized tryptophan-binding aptamer. A) A small area of the family C aptamer matches the consensus sequence for a previously characterized tryptophan aptamer **B**) described by Majerfeld et al., 2010.

that it was, indeed, non-selective for 5HW over tryptophan, and that its K_D was similar to the previously described tryptophan aptamer. It bound 5HW with a K_D of 38 ± 14 µM, while it bound tryptophan with a K_D of 20 ± 5 µM, while the previous study reported a K_D of 15 µM for the tryptophan aptamer, which is within error of the obtained K_D for the family C aptamer. Thus, the structure is an attractive target for crystallization, not only because it can be compared to the family B structure, but also because it may help describe how different RNA sequences can bind the same ligand with completely different affinities and specificities.

A   **B**
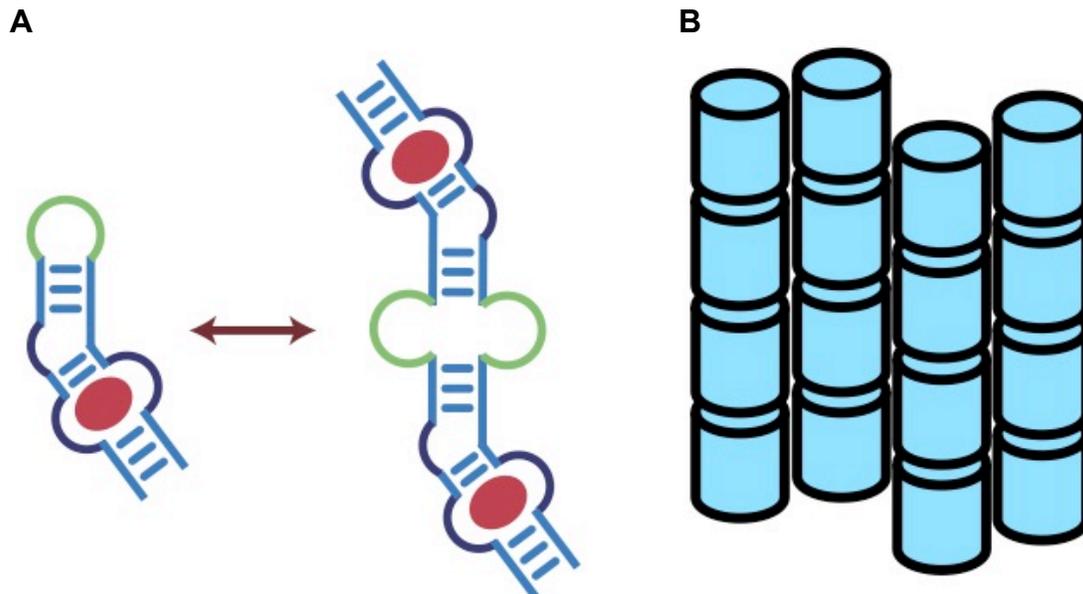
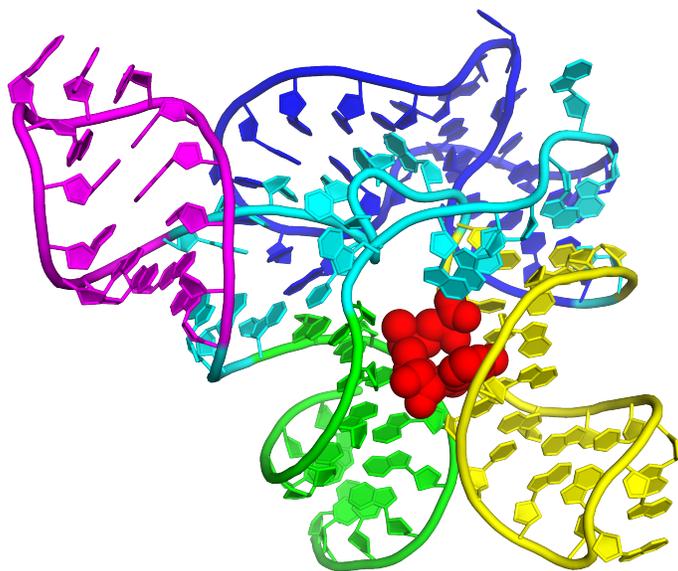
Figure 17: Difficulties in crystallizing small hairpins. A) Small hairpins are self-complementary and can thus form double helices. Folding heterogeneity can lead to lattice poisoning. **B**) Helices can stack end-on-end to form long fibers, but they have few crystal contacts along their edges. Thus, the unit cell can be composed of non-repeating units.

The family C aptamer, hereupon referred to as the tryptophan aptamer, presents many difficulties for crystallization. Firstly, the structure of a small hairpin can result in folding heterogeneity. A small hairpin is self-complementary, so it can readily dimerize into a double helix (Figure 17). Thus, at any given time, a part of the RNA population may be in a double helix, while a different part of the population may be folded into the single-stranded hairpin. Pool heterogeneity is detrimental to crystal formation because if a double stranded RNA packs into the crystal lattice of RNAs that are all single-stranded hairpins or vice versa, the crystal lattice will be poisoned, and no uniform crystal can form.

Secondly, even if the RNA folds could be made uniform, double helices and small hairpins still face problems in long fiber formation (Figure 17). Helices lack nucleotides that can form crystal contacts along their backbones, while most of their potential for forming crystal contacts is at the ends of the helices. Thus, the helices can easily stack one on top of another other utilizing end-to-end

crystal contacts to form long fibers. However, each fiber cannot communicate

effectively with neighboring fibers because of the lack of crystal contacts. As a result, fibers will be oriented against each other randomly, and although a crystal lattice can form and electron density may even be calculated, the electron density will show only a long barrel of homogeneous electron density, as

Figure 18: SAM-I Riboswitch. The P4 (magenta) and P3 (yellow) helices extend into space, so it was reasoned that they could be elongated without perturbing SAM binding (Montange et al., 2009).

each fiber is not oriented uniformly in the unit cell.

Thus, in order to crystallize the hairpin, the S-adenosylmethionine-I (SAM-I) riboswitch was used as a scaffold to which the tryptophan aptamer was appended. Although previous studies have utilized protein/RNA hybrid scaffolds to aid in crystallization (Ferré-D'Amaré, 2010), no study to-date has used a riboswitch as a scaffold for crystallization. Members of the Batey laboratory previously solved the SAM-I riboswitch structure to high resolution (Montagne et al., 2006). The structure has two helices, P3 and P4, which extend into space (Figure 18), and it was reasoned that the tryptophan aptamer could be affixed onto these helices without perturbing the SAM-I structure and allowing for tryptophan binding. Since the SAM-I riboswitch is not self-complementary and will not form long helix fibers, this approach could solve problems crystallizing the small hairpin aptamer.
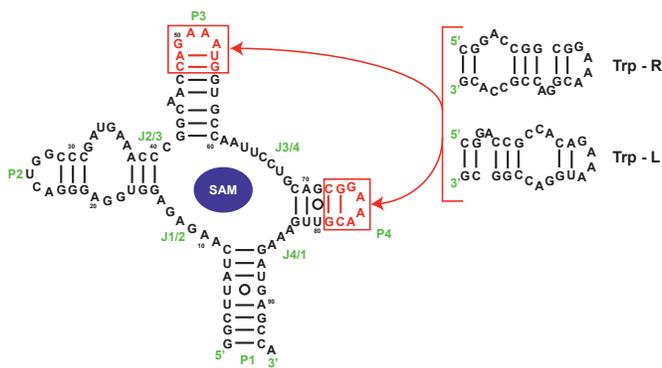
4.2 Designing SAM-I Tryptophan Fusion Sequences that Bind Tryptophan

In order to prove that the tryptophan aptamer fused to the SAM-I riboswitch bound tryptophan using the same mechanism as the tryptophan aptamer alone, ITC was employed. Generally, if an aptamer has the same K_D and sequence as another aptamer, it can be inferred that the binding mechanism is the same. Thus, the K_D's of each fusion construct were measured before crystallization was attempted so that any resulting structures would be sure to employ the same tryptophan binding mechanism as the free tryptophan aptamer.

The first sequences constructed were designed as the minimal constructs believed to be required for the RNA to maintain SAM binding while introducing tryptophan binding (Figure 19). Minimal constructs are desirable in crystallizations because nonessential elements may increase the degrees of freedom in the structure, which in turn may decrease the resolution limit. Thus, only two base pairs were added between the tryptophan aptamer and the last residues critical for SAM binding in the P3 and P4 helices. Because the orientation of a capping loop on the tryptophan aptamer is not conserved, the

Figure 19: First SAM-tryptophan fusion constructs. Constructs minimizing linker length between the SAM binding region and the tryptophan aptamer were designed. 4 constructs resulted, where each Trp-R and –L were affixed to P3 and P4, respectively.

aptamer was appended in both possible orientations, where the capping loop was on opposite ends. These two orientations were termed Trp-L and Trp-R. The GAAA capping tetraloop was chosen for its potential to form crystal contacts with neighboring molecules. By putting each tryptophan orientation onto each the P3 and P4 helices respectively, four constructs were designed.

The constructs were PCR amplified, transcribed, and purified using the same protocol described in RNA preparation for family B crystallization. ITC was

performed under conditions previously employed for measuring the K_D of the tryptophan aptamer, using a buffer containing 50 mM sodium HEPES pH 7.0, 250 mM sodium chloride, 5 mM magnesium chloride, and 5 mM calcium chloride (Majerfeld et al., 2010). Each construct was measured for tryptophan binding using ITC, but none could effectively bind tryptophan in either the presence or absence of SAM. Thus, new constructs were created to potentially remedy the lack of tryptophan binding.

The second iteration of constructs augmented the helix length between the tryptophan aptamer and the SAM-I binding pocket from two base pairs to four (Figure 20). Additionally, the Trp-R orientation of the tryptophan aptamer was not tested to reduce iterations required. However, an additional base pair was added to the Trp-L sequence
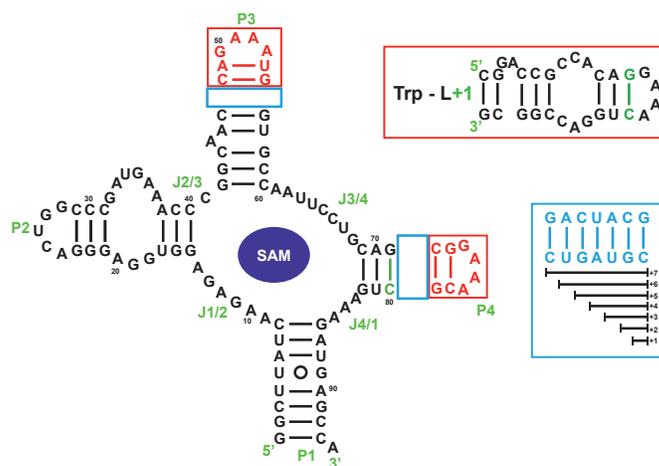
Figure 20: SAM-I tryptophan fusion constructs with increased helix length. The helix linker between the tryptophan aptamer was increased from 2 base pairs a minimum of 4 base pairs and a maximum of 9 base pairs. Additionally, one base pair was added to the end of the Trp-L construct to increase helix stability.

between the aptamer and the capping tetraloop to ensure helix stability. Thus, two constructs were designed: SAMTrp-L+1 P3+2 (appending the Trp-L aptamer orientation with the addition of one base pair to the P3 helix with two additional base pairs in the adjoining region) and SAMTrp-L+1 P4+2 (the equivalent design appended in the P4 helix instead of the P3 helix).

Again, these constructs were PCR amplified, transcribed, and purified. They were tested for binding via ITC in the same buffer conditions, and tryptophan binding was detected. The K_D of the SAMTrp-L+1 P3+2 construct was

determined to be 16 ± 2 µM, and the K_D of the SAMTrp-L+1 P4+2 construct was determined to be 16 ± 1 µM. Both values are within acceptable error of values obtained previously for the tryptophan aptamer, which ranged from 12 to 15 µM (Majerfeld et al., 2010). Thus, it can be strongly suggested that the tryptophan aptamer bound in the same manner in both the free aptamer and the SAM-I fusions. Additionally, it was reasoned that mutants with longer linker helix lengths between SAM-I and the Trp-L+1 aptamer sequence would also bind tryptophan with similar K_D's. Therefore, several SAM-I tryptophan fusion sequences with helix linker lengths +2 – +7 were screened for crystallization (Figure 20).

4.3 Crystallizing SAM-I Tryptophan Fusion RNAs

 SAM-I tryptophan fusion constructs using the Trp-L+1 aptamer and helix linker lengths +2 – +7 were PCR amplified, transcribed, and purified. Hanging-drop crystallization was utilized as formerly described. Crystal screens were again employed to search for initial mother liquor conditions that could encourage crystallization. The Hampton Research Nucleic Acid Mini Screen, Natrix 1 Screen, Natrix 2 Screen, and Crystal Screen were all tested, in addition to the Qiagen Nucleix Suite and PEG II's Suite for a total of 360 unique mother liquor conditions. All screens were used as described in protocols provided by each

A　　　　　　　　　　**B**
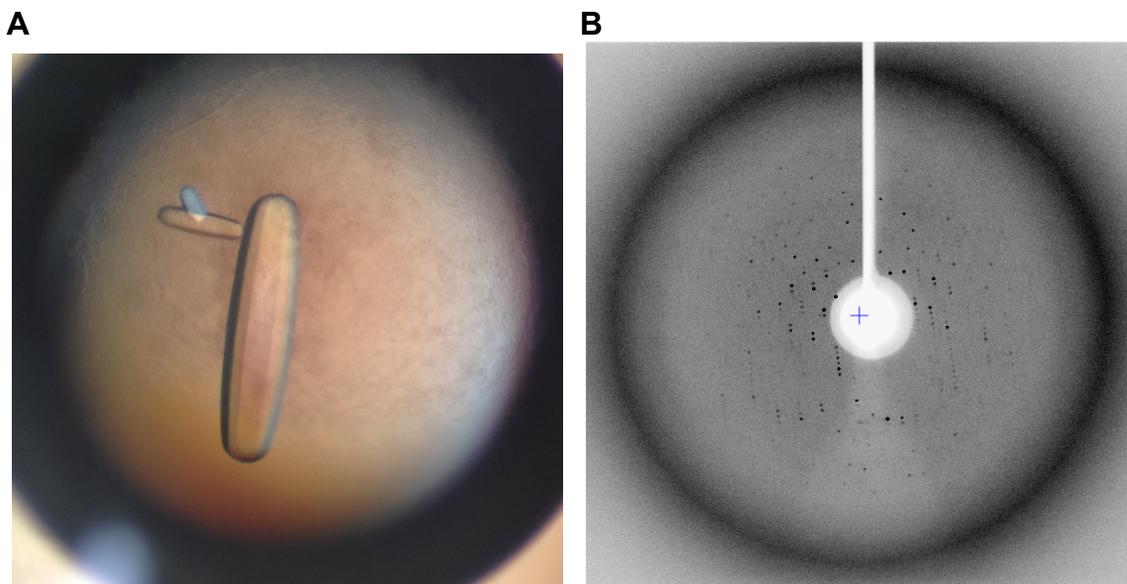
Figure 21: SAMTrp-L+1 P4+5 crystals. A) Large SAMTrp-L+1 P4+5 crystals grow reproducibly when barium chloride is added to the mother liquor. **B**) However, the crystals only diffract to a resolution of ~5.0 Å.

company, using RNA solutions 400 µM in RNA, 2 mM in SAM, and 2 mM in tryptophan in .5X Tris-EDTA buffer.

SAMTrp-L+1 P4+5 was able to crystallize in several different Natrix 2 conditions. Mother liquor conditions were optimized, and the best crystals were observed in 1 mM cobalt hexammine, 9% ethanol, 50 mM sodium cacodylate pH 6.5, 45 – 55 mM magnesium acetate, and 100 – 110 mM barium chloride. Crystals were cryoprotected with a 10 µL of a solution isotonic to the above mother liquor with an additional 15% 2-methyl-2,4-pentanediol (MPD) and flash frozen in liquid nitrogen. Several crystals were screened, but even the best crystals could not achieve below 5.0 Å resolution (Figure 21). Generally, around 3.5 Å is the maximum resolution limit required to accurately determine a structure. Thus, it became apparent that the structure could not be solved to high resolution with the SAMTrp-L+1 P4+5 construct.

Currently, a new set of mutations to the SAMTrp-L+1 P4+5 construct is being investigated. Three mutations were chosen for their success in increasing resolution on the original SAM-I riboswitch: an A14 deletion, a U34 to C mutation, and an A94 to G mutation (Montagne et al., 2009). The mutations occur on residues that are oriented out from the riboswitch such that they can potentially form crystal contacts with neighboring RNA molecules (Figure 22). Three constructs containing one mutation each were created and are presently being tested for a potential increase in resolution.

A   **B**
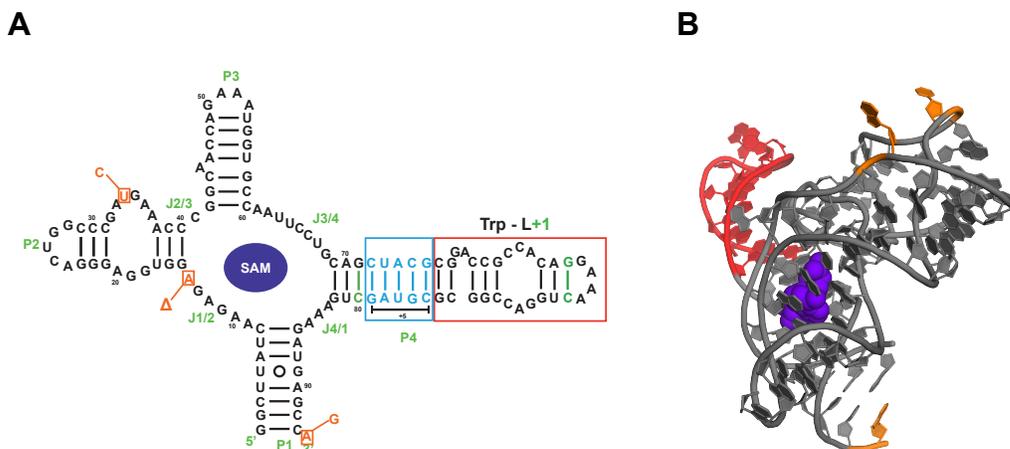
Figure 22: SAMTrp-L+1 P4+5 mutants. A) Three different mutants will be screened for their ability to form high-resolution crystals. **B**) The mutations occur on residues extending from the surface of the RNA, so they may affect crystal contacts with neighboring RNA molecules (Montagne et al., 2009).

5. Discussion

Although riboswitches are attractive targets for synthetic biologists, past research has generally only been able to broadly utilize natural riboswitch aptamer domains to create a response. This study has presented a possible mechanism by which an aptamer might retain regulatory capabilities while recognizing a novel ligand. The purine riboswitch was successfully reselected for binding to 5HW. The crystal structure of the new, 5HW-sensing aptamer demonstrated the ability of an RNA to bind a novel ligand while maintaining tertiary architectures necessary for gene regulation. This approach may prove effective in creating a synthetic riboswitch system, which Ely is currently attempting with the family B aptamer. If successful, synthetic riboswitches could be used for a range of applications, from ligand sensing to feedback regulation in fully synthetic pathways.

Furthermore, determining the structure of the tryptophan aptamer will increase our understanding of how different RNAs can bind the same ligand. The tryptophan aptamer has a different binding affinity and specificity as compared to the family B aptamer, despite arising from the same selection and binding the same ligand. This suggests that it will likely employ a completely different binding mechanism as compared to the family B aptamer. Investigating these different binding mechanisms may provide insight into RNA-ligand interactions. In addition, if successful, the method employed to crystallize the small hairpin tryptophan aptamer will increase our toolbox for determining the structure of small, difficult to crystallize RNAs. Thus, the structure of the tryptophan aptamer will not only deepen our understanding of RNA ligand recognition, but it will present a new tool for RNA crystallographers to potentially crystallize difficult RNAs.

6. Appendices

Oligomer Sequences

Construct	Oligomers
Randomized purine library	**xpt-pbuX scaffold (bold denotes the first transcribed nucleotide):** GCGCGCGAATTCTAATACGACTCACTATA**G**GACTTCGGTCCAAGCTAATGCACTCNNNNNNNCGCGTGGATATGGCACGCANNNNNNNNNNGGGCACCGTAAATGTCCNNNNNNNGGGTGCATTAGCAAAATCGGGCTTCGGTCCGGTTC **T7 Site Appending Primer *xpt*:** GCGCGCGAATTCTAATACGACTCACTATAGGACTTCGGTCCAAGCTAATGCACTC **RT-PCR Primer *xpt*:** 5'-GAACCGGACCGAAGCCCG

Construct	Oligomers
x17 P1-5	**x17p1-5 gene:** GCGCGCGAATTCTAATACGACTCACTATAGGATCTGATGATCGCGTGGATATGGCACGCATTGAATTGTTGGACACCGTAAATGTCCTAACACGGTCC **5' x17p1-5 Gen:** GCGCGCGAATTCTAATACGACTCACTATAGGATCTGATGATCGCGTGG **3' x17p1-5m:** mUmGGACCGTGTTAGGACATTTACGG
x17 P1-6	**x17p1-6 gene:** GCGCGCGAATTCTAATACGACTCACTATAGGACACTGATGATCGCGTGGATATGGCACGCATTGAATTGTTGGACACCGTAAATGTCCTAACACGTGTCC **5' x17p1-6 Gen:** GCGCGCGAATTCTAATACGACTCACTATAGGACACTGATGATCGCGTGG **3' x17p1-6m:** mUmGGACACGTGTTAGGACATTTACGG
x17 P1-7	**x17p1-7 gene:** TAATACGACTCACTATAGGACATCTGATGATCGCGTGGATATGGCACGCATTGAATTGTTGGACACCGTAAATGTCCTAACACGGTGTCC **5' Gen:** GCGCGCGAATTCTAATACGACTCACTATAGG **3' x17p1-7:** GGACACCGTGTTAGGACATTTACGG **3' x17p1-7m:** mUmGGACACCGTGTTAGGACATTTACGG
x17 P1-8	**x17p1-8 gene:** GCGCGCGAATTCTAATACGACTCACTATAGGACACACTGATGATCGCGTGGATATGGCACGCATTGAATTGTTGGACACCGTAAATGTCCTAACACGTGTGTCC **5' x17p1-8 Gen:** GCGCGCGAATTCTAATACGACTCACTATAGGACACACTGATGATCGCGTGG **3' x17p1-8m:** mUmGGACACACGTGTTAGGACATTTACGG

Construct	Oligomers (If sequence not listed, it is the same oligo as cells above)
SAMTrp-R P3	**F T7.Saml.xtal.01**: GCGCGCGAATTCTAATACGACTCACTATAGGCTTATCAAGAGAGGTGGAG **R Saml.p2.wt.02**: GTTGCCGGGTTTCATCGGGCCAGTCCCTCCACCTCTCTTGATAAGC **F Saml.trpR.p3.03**: CGATGAAACCCGGCAACCGGACCGGCGGAAACGGACCGCCACGGTGCCAATTCCTGCAGC **R Saml.p4.wt.04**: mUmGGCTCATCTTTCAACGTTTCCGCTGCAGGAATTGGCAC
SAMTrp-L P3	**F T7.Saml.xtal.01** **R Saml.p2.wt.02** **F Saml.trpL.p3.03**: CGATGAAACCCGGCAACCGGACCGCCACTGAAAAGGACCGGCGGTGCCAATT

	CCTGCAGC **R SamI.p4.wt.04**
SAMTrp- L P4	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.p3.wt.03**: CGATGAAACCCGGCAACCGAAAGGTGCCAATTCCTGCAGC **R SamI.trpL.p4.04:** mUmGGCTCATCTTTCAACGCCGGTCCTTTTCAGTGGCGGTCCGCTGCAGGAAT TGGCAC
SAMTrp- L P4	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.p3.wt.03** **R SamI.trpR.p4.04**: mUmGGCTCATCTTTCAACGTGGCGGTCCGTTTCCGCCGGTCCGCTGCAGGAAT TGGCAC **R SamI.p4.wt.04**
SAMTrp- L+1 P3+2	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** F SamI.trpL+1.p3+2.03 CGATGAAACCCGGCAACCGCGGACCGCCACTGGAAACAGGACCGGCGCGGTG CCAATTCCTGCAGC **R SamI.p4.wt.04**
SAMTrp- L+1 P4+2	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.p3.wt.03** **R SamI.trpL+1.p4+2.04** mUmGGCTCATCTTTCAGCGCGCCGGTCCTGTTTCCAGTGGCGGTCCGCGCTG CAGGAATTGGCAC
SAMTrp- L+1 P3+3	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.trpL+1.p3+3.03**: CGATGAAACCCGGCAACACGCGGACCGCCACTGGAAACAGGACCGGCGCGTG TGCCAATTCCTGCAGC **R SamI.p4.wt.04**
SAMTrp- L+1 P4+3	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.p3.wt.03** **R SamI.trpL+1.p4+3.04** mUmGGCTCATCTTTCAGACGCGCCGGTCCTGTTTCCAGTGGCGGTCCGCGTCT GCAGGAATTGGCAC
SAMTrp- L+1 P3+4	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.trpL+1.p3+4.03** CGATGAAACCCGGCAACTACGCGGACCGCCACTGGAAACAGGACCGGCGCGT AGTGCCAATTCCTGCAGC **R SamI.p4.wt.04**
SAMTrp- L+1 P4+4	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.p3.wt.03** **R SamI.trpL+1.p4+4.04** mUmGGCTCATCTTTCAGTACGCGCCGGTCCTGTTTCCAGTGGCGGTCCGCGTA CTGCAGGAATTGGCAC
SAMTrp- L+1 P3+5	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.trpL+1.p3+5.03** CGATGAAACCCGGCAACCTACGCGGACCGCCACTGGAAACAGGACCGGCGCG

	TAGGTGCCAATTCCTGCAGC
	R SamI.p4.wt.04
SAMTrp-L+1 P4+5	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.p3.wt.03** **R SamI.trpL+1.p4+5.04** mUmGGCTCATCTTTCAGCTACGCGCCGGTCCTGTTTCCAGTGGCGGTCCGCGT AGCTGCAGGAATTGGCAC
SAMTrp-L+1 P3+6	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.trpL+1.p3+6.03** CGATGAAACCCGGCAACACTACGCGGACCGCCACTGGAAACAGGACCGGCGC GTAGTGTGCCAATTCCTGCAGC **R SamI.p4.wt.04**
SAMTrp-L+1 P4+6	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.p3.wt.03** **R SamI.trpL+1.p4+6.04** mUmGGCTCATCTTTCAGACTACGCGCCGGTCCTGTTTCCAGTGGCGGTCCGCG TAGTCTGCAGGAATTGGCAC
SAMTrp-L+1 P3+7	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.trpL+1.p3+7.03** CGATGAAACCCGGCAACGACTACGCGGACCGCCACTGGAAACAGGACCGGCG CGTAGTCGTGCCAATTCCTGCAGC **R SamI.p4.wt.04**
SAMTrp-L+1 P4+7	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.p3.wt.03** **R SamI.trpL+1.p4+7.04** mUmGGCTCATCTTTCAGGACTACGCGCCGGTCCTGTTTCCAGTGGCGGTCCGC GTAGTCCTGCAGGAATTGGCAC
SAMTrp-L+1 P4+5 dA14	**F T7.SamI.xtal.01.dA14** GCGCGCGAATTCTAATACGACTCACTATAGGCTTATCAAGAGGGTGGAG **R SamI.p2.02.dA14** GTTGCCGGGTTTCATCGGGCCAGTCCCTCCACCCTCTTGATAAGC **F SamI.p3.wt.03** **R SamI.trpL+1.p4+5.04**
SAMTrp-L+1 P4+5 U34C	**F T7.SamI.xtal.01** **R SamI.p2.02.U34C** GTTGCCGGGTTTCGTCGGGCCAGTCCCTCCACCTCTCTTGATAAGC **F SamI.p3.03.U34C.P3wt+1** CGACGAAACCCGGCAACCAGAAATGGTGCCAATTCCTGCAGC **R SamI.trpL+1.p4+5.04**
SAMTrp-L+1 P4+5 mCm	**F T7.SamI.xtal.01** **R SamI.p2.wt.02** **F SamI.p3.wt.03** **R SamI.trpL+1.p4+5.04.mCm** mCmGGCTCATCTTTCAGCTACGCGCCGGTCCTGTTTCCAGTGGCGGTCCGCGT AGCTGCAGGAATTGGCAC

References

Allner O, Nilsson L, Villa A. 2013. Loop-loop interaction in an adenine-sensing riboswitch: A molecular dynamics study. RNA: A Publication of the RNA Society 19(7):916-26.

Batey RT. 2012. Structure and mechanism of purine-binding riboswitches. Q Rev Biophys. 45(3):345-81.

Batey RT, Gilbert SD, Montange RK. 2004. Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. Nature 432:411-5.

Berens C, Thain A, Schroeder R. 2001. A tetracycline-binding RNA aptamer. Bioorg. Med. Chem. 9(10):2549-56.

Bocobza S, Adato A, Mandel T, Shapira M, Nudler E, Aharon A. 2007. Riboswitch-dependent gene regulation and its evolution in the plant kingdom. Genes Dev 21(22):2874-9.

Ceres P, Trausch JJ, Batey RT. 2013. Engineering modular 'ON' RNA switches using biological components. Nucleic Acids Research .

Ceres P, Garst AD, Marcano-Velazquez JG, Batey RT. 2013. Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. Acs Synthetic Biology 2(8):463-72.

Cheah MT, Wachter A, Sudarsan N, Breaker RR. 2007. Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. Nature 447(7143):497-U7.

Chen AGY, Sudarsan N, Breaker RR. 2011. Mechanism for gene control by a natural allosteric group I ribozyme. RNA-Publ RNA Soc 17(11):1967-72.

Collins JA, Irnov I, Baker S, Winkler WC. 2007. Mechanism of mRNA destabilization by the glmS ribozyme. Genes Dev 21(24):3356-68.

Croft MT, Moulin M, Webb ME, Smith AG. 2007. Thiamine biosynthesis in algae is regulated by riboswitches. Proc Natl Acad Sci U S A 104(52):20770-5.

Du H, Yakhnin AV, Dharmaran S, Babitzke P. 2000. *trp* RNA-Binding Attenuation Protein-5′ Stem-Loop RNA Interaction Is Required for Proper Transcription Attenuation Control of the *Bacillus subtilis trpEDCFBA*Operon. J. Bacteriol. 182(7):1819-1827 .

Edwards TE and Ferre-D'Amare AR. 2006. Crystal structures of the thi-box riboswitch bound to thiamine pyrophosphate analogs reveal adaptive RNA-small molecule recognition. Structure 14(9):1459-68.

Ellington A and Szostak J. 1990. In vitro selection of RNA molecules that bind specific ligands. Nature 346(6287):818-22.

Ferré-D'Amaré AR. 2010. Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. Methods 52(2):159-167.

Garst AD and Batey RT. 2009. A switch in time: Detailing the life of a riboswitch. Biochimica Et Biophysica Acta (BBA) - Gene Regulatory Mechanisms 1789(9–10):584-91.

Garst AD, Porter EB, Batey RT. 2012. Insights into the regulatory landscape of the lysine riboswitch. J Mol Biol 423(1):17-33.

Garst AD, Edwards AL, Batey RT. 2011. Riboswitches: Structures and mechanisms. Cold Spring Harbor Perspect Biol 3(6):a003533.

Garst AD, Heroux A, Rambo RP, Batey RT. 2008. Crystal structure of the lysine riboswitch regulatory mRNA element. J Biol Chem 283(33):22347-51.

Gilbert SD, Rambo RP, Van Tyne D, Batey RT. 2008. Structure of the SAM-II riboswitch bound to S-adenosylmethionine. Nat Struct Mol Biol 15(2):177-82.

Gilbert SD, Reyes FE, Batey RT. 2009. Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. Structure 17:857-868

Gilbert SD, Stoddard CD, Wise SJ, Batey RT. 2006. Thermodynamic and kinetic characterization of ligand binding to the purine riboswitch aptamer domain. J Mol Biol 359(3):754-68.

Haller A, Souliere MF, Micura R. 2011. The dynamic nature of RNA as key to understanding riboswitch mechanisms. Acc Chem Res 44(12):1339-48.

Holbrook SR. 2008. Structural principles from large RNAs. Ann Rev Biophys 37:445-64.

Hollands K, Proshkin S, Sklyarova S, Epshtein V, Mironov A, Nudler E, Groisman EA. 2012. Riboswitch control of rho-dependent transcription termination. Proc Natl Acad Sci U S A 109(14):5376-81.

Huang L, Serganov A, Patel DJ. 2010. Structural insights into ligand recognition by a sensing domain of the cooperative glycine riboswitch. Mol Cell 40(5):774-86.

Huang L, Ishibe-Murakami S, Patel DJ, Serganov A. 2011. Long-range pseudoknot interactions dictate the regulatory response in the tetrahydrofolate riboswitch. Proc Natl Acad Sci U S A 108(36):14801-6.

Irnov I, Sharma CM, Vogel J, Winkler WC. 2010. Identification of regulatory RNAs in bacillus subtilis. Nucleic Acids Res 38(19):6637-51.

Johnson JE, Jr., Reyes FE, Polaski JT, Batey RT. 2012. B-12 cofactors directly stabilize an mRNA regulatory switch. Nature 492(7427):133,+.

Keasling JD. 2012. Synthetic biology and the development of tools for metabolic engineering. Metabolic Engineering 14(3):189-195.

Krasilnikov AS, Mondragon A. 2003. On the occurrence of the T-loop RNA folding motif in large RNA molecules. RNA. 9(6):640-3.

Livny J and Waldor MK. 2010. Mining regulatory 5'UTRs from cDNA deep sequencing datasets. Nucleic Acids Research 38(5):1504-14.

Loh E, Dussurget O, Gripenland J, Vaitkevicius K, Tiensuu T, Mandin P, Repoila F, Buchrieser C, Cossart P, Johansson J. 2009. A trans-acting riboswitch controls expression of the virulence regulator PrfA in listeria monocytogenes. Cell 139(4):770-9.

Lu C, Smith AM, Fuchs RT, Ding F, Rajashankar K, Henkin TM, Ke A. 2008. Crystal structures of the SAM-III/S-MK riboswitch reveal the SAM-dependent translation inhibition mechanism. Nat Struct Mol Biol 15(10):1076-83.

Luo GX, Taylor J. Template switching by reverse transcriptase during DNA synthesis. J Virol. 64(9): 4321–4328.

Montange RK, Batey RT. 2008. Riboswitches: emerging themes in RNA structure and function. Annu Rev Biophys. 37:117-33.

Montange RK, Mondragon E, van Tyne D, Garst AD, Ceres P, Batey RT. 2010. Discrimination between closely related cellular metabolites by the SAM-I riboswitch. J. Mol. Biol. 396(3):761-72.

Piasecki, Shawn K, Bradley Hall, and Andrew D. Ellington. "Nucleic Acid Pool Preparation and Characterization." Nucleic Acid and Peptide Aptamers: Methods and Protocols. Ed. Gunter Mayer. New York: Humana Press, 2009. 3-18.

Raghavan R, Groisman EA, Ochman H. 2011. Genome-wide detection of novel regulatory RNAs in E. coli. Genome Res 21(9):1487-97.

Ren A, Rajashankar KR, Patel DJ. 2012. Fluoride ion encapsulation by Mg2+ ions and phosphates in a fluoride riboswitch. Nature 486(7401):85-U1501.

Rieder R, Lang K, Graber D, Micura R. 2007. Ligand-induced folding of the adenosine deaminase A-riboswitch and implications on riboswitch translational control. Chembiochem 8(8):896-902.

Serganov A, Polonskaia A, Phan A, Breaker R, Patel D. 2006. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. Nature 441(7097):1167-71.

Serganov A, Yuan Y, Pikovskaya O, Polonskaia A, Malinina L, Phan A, Hobartner C, Micura R, Breaker R, Patel D. 2004. Structural basis for discriminative regulation of gene expression by adenine- and guanine-sensing mRNAs. Chem Biol 11(12):1729-41.

Serganov A and Nudler E. 2013. A decade of riboswitches. Cell 152(1–2):17-24.

Serganov A, Huang L, Patel DJ. 2008. Structural insights into amino acid binding and gene control by a lysine riboswitch. Nature 455(7217):1263-U76.

Sinha J, Reyes SJ, Gallivan JP. 2010. Reprogramming bacteria to seek and destroy an herbicide. Nat Chem Biol 6(6):464-70.

Stoddard CD, Montange RK, Hennelly SP, Rambo RP, Sanbonmatsu KY, Batey RT. 2010. Free state conformational sampling of the SAM-I riboswitch aptamer domain. Structure 18(7):787-97.

Thore S, Leibundgut M, Ban N. 2006. Structure of the eukaryotic thiamine pyrophosphate riboswitch with its regulatory ligand. Science 312(5777):1208-11.

Trausch JJ, Ceres P, Reyes FE, Batey RT. 2011. The structure of a tetrahydrofolate-sensing riboswitch reveals two ligand binding sites in a single aptamer. Structure 19(10):1413-23.

Tuerk C and Gold L. 1990. Systematic evolution of ligands by exponential enrichment - rna ligands to bacteriophage-T4 dna-polymerase. Science 249(4968):505-10.

Wachter A, Tunc-Ozdemir M, Grove BC, Green PJ, Shintani DK, Breaker RR. 2007. Riboswitch control of gene expression in plants by splicing and alternative 3 ' end processing of mRNAs. Plant Cell 19(11):3437-50.

Wickiser J, Winkler W, Breaker R, Crothers D. 2005. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. Mol Cell 18(1):49-60.

Wickiser J, Cheah M, Breaker R, Crothers D. 2005. The kinetics of ligand binding by an adenine-sensing riboswitch. Biochemistry 44(40):13404-14.

Wilkinson KA, Merino EJ, Weeks KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nature Protocols 1(3):1610-1616.

Winkler W, Nahvi A, Breaker R. 2002. Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. Nature 419(6910):952-6.

Winkler W and Breaker R. 2005. Regulation of bacterial gene expression by riboswitches. Annu Rev Microbiol 59:487-517.

Acknowledgments

      Firstly, I would like to thank Dr. Robert Batey for providing me with the opportunity to pursue research in the laboratory, including providing financial sponsorship so that I could continue to learn over the summers. The training I've gained working in the laboratory has greatly enriched my undergraduate education and played the major role in my decision to go to graduate school next year, which I never even considered prior to joining the lab. I would also like to thank my thesis committee members, Dr. Joe Falke and Dr. Nancy Guild, for generously offering their time to advise my thesis defense. Additionally, I would like to give a very special thanks to Ely Porter, whose seemingly infinite patience is nothing short of extraordinary; I couldn't have been luckier to have him as a mentor. His guidance has been invaluable to my experience in the laboratory and beyond over the past three years, and I will be forever grateful for his mentorship. I would also like to thank all of the other Batey lab members and alumni, who have each provided something unique to add to my experiences in the laboratory. I am grateful for the generous support provided by the CU Undergraduate Research Opportunities program and the Howard Hughes Medical Institute, which has provided me with financial support over the past two semesters. Finally, I would like to thank my friends and family, who have never been anything but accommodating of my constantly busy lifestyle. The support provided by each of the aforementioned parties has allowed me to grow and learn throughout my undergraduate career in ways I never could have alone, so thank you!