Effect of Variation in an Essential Folding Element on the Cobalamin Riboswitch

Marcus A. Urquijo

Defense Date: April 3, 2018

Department of Chemistry/Biochemistry

April 3, 2018

Thesis Advisor: Robert T. Batey (CHEM/BIOCHEM)

Honors Council Representative: Jeffery Cameron (CHEM/BIOCHEM)

Committee: Robert Batey (CHEM/BIOCHEM), Jeffery Cameron (CHEM/BIOCHEM), Shelley Copley (MCDB)

Table of Contents

Abstract

Riboswitches are RNA-based genetic regulatory elements that control gene expression without the need for protein cofactors. These RNA motifs are found within the 5’ untranslated regions of mRNA and promote genetic regulation by altering the structure of RNA through binding small molecule effector ligands that cause conformational changes that repress or activate translation/transcription. The env8 cobalamin riboswitch binds cyanocobalamin (CNCbl) to repress gene expression of the downstream mRNA. It accomplishes this through an interplay of two separate domains; an aptamer domain which binds the effector ligand and a regulatory domain which contains the genetic switch. Within the aptamer domain, there are several important constructs that promote the tertiary structure of the cobalamin binding pocket as well as the T-Loop that interacts with the regulatory domain. This study is aimed to observe the effect of variation within the L4-L6 T-Loop interaction that organizes the structure of the cobalamin binding pocket.

It was observed that a highly conserved consensus sequence is required for maintenance of riboswitch functionality. Specifically, nucleotides 30 and 33 on the L4 in addition to nucleotide 61 on the L6 are highly conserved and therefore must play a key role in the formation of the binding pocket. This region of the riboswitch was seen to be conserved regardless of the efficiency of the riboswitch and can be assumed to have critical importance in overall functionality.

Chapter 1: Introduction

Gene Regulation

Expression of the numerous genes in a cell must be tailored to provide the appropriate levels of RNA and protein production at all times. Gene regulation systems, therefore, must have the ability to respond precisely to specific signals, rapidly bring about their intended genetic effect, and have sufficient dynamic character to fine-tune the level of expression for hundreds of different genes[1]. Cells can control gene expression through a number of methods and mechanisms; each pertaining to a unique pathway in which it recognizes an effector ligand and produces the needed regulatory response. Within these levels, numerous factors typically work together to influence transcription, translation, mRNA processing/degradation, and other mechanisms that control levels of gene products in the cell[2].

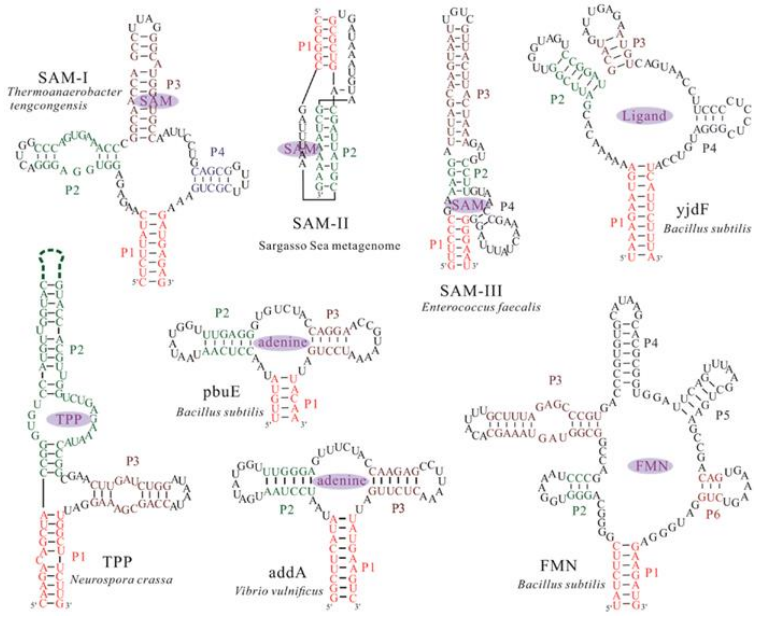

Figure 1 (Gong, S., Wang, Y., Wang, Z., & Zhang, W. (2017)): A depiction of aptamer domains isolated from distinct types/classes of riboswitches. The secondary structure of the aptamer domain is shown here along with ligands of interested located within the binding sites. This figure does not show the downstream regulatory domain.

Originally when the central dogma was first conceived, regulation of gene expression was believed to be controlled primarily by protein repressors and activators. However, the discovery of the transcriptional attenuation by Charles Yonofsky in 2000, gave rise to the use of RNA in gene control systems. Yonofsky's work investigated RNA attenuation which regulated the tryptophan operon. RNA generates two separate hairpin loops, a terminator and anti-terminator. These secondary structures in conjunction with the ribosomal complex regulate the synthesis of tryptophan[23]. This discovery changed the perceived role of RNA in that it plays a critical role in genetic regulation. In addition to RNA-protein regulatory complexes, metabolite-sensing RNAs found in mRNA are capable of modulating gene expression upon binding of an effector ligand. The discovery of RNA's ability to regulate gene expression in a protein independent manner led to the research into regulatory RNAs termed "riboswitches". These "riboswitches" form within the leader sequences of messenger RNAs and bind their target metabolite with the affinity and specificity required for the precise regulation of gene expression[4].

Riboswitches

These regulatory RNAs are autonomous noncoding RNA elements that monitor the cellular environment and control gene expression[1,7,8,9]. These segments of mRNA bind selectively to small molecule effector ligands and alter gene expression at the translational or transcriptional level. More than 40 classes of riboswitches that respond to changes in the presence of specific small molecule ligands ranging anywhere from amino acids to coenzymes are currently known[6]. Riboswitches are a common regulatory module used in bacteria for genetic regulation at the transcriptional and translational level. The distribution of classes of riboswitches within the bacteria kingdom was

characterized by McCown and Phillip John in 2014[24]. The results of the study concluded that riboswitches are a widely distributed regulatory modulate that is included in almost all forms of bacteria. Riboswitches such as the cobalamin riboswitch, TPP, and SAM-I are some examples of widely distributed riboswitches (Figure 2).

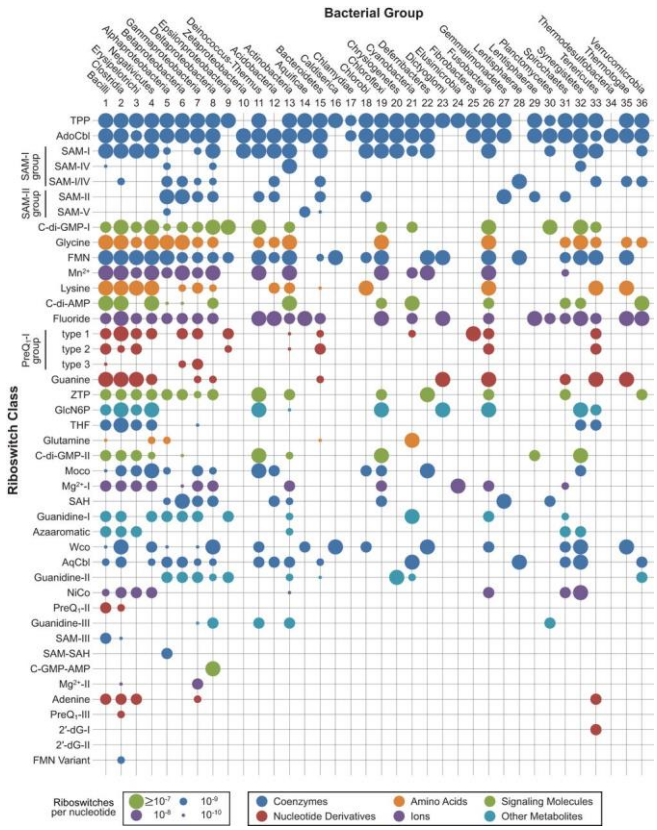

Figure 2 (McCown et al): A diagram depicting riboswitches contained within diverse types of bacteria. TPP, AdoCbl, and SAM-I are included within most bacteria. It can be shown that all bacteria contain a riboswitch in some form

Riboswitches regulatory function is achieved through an interplay of two domains: a receptor (aptamer) domain and a regulatory domain. The aptamer domain binds the effector ligand of interest and elicits an interaction with the downstream regulatory domain (Figure 1)[4]. Regulatory domains contain the structural switch to regulate gene expression; this can be in the form of sequestering the ribosomal binding site (RBS) or recruiting other regulatory modules[4]. These effector ligands can range from nucleotides such as adenine, to complex ligands such as cobalamin B-12 (Figure 1). The wide variety of ligands corresponds with the requirement of aptamer domains being capable of sifting through the plethora of

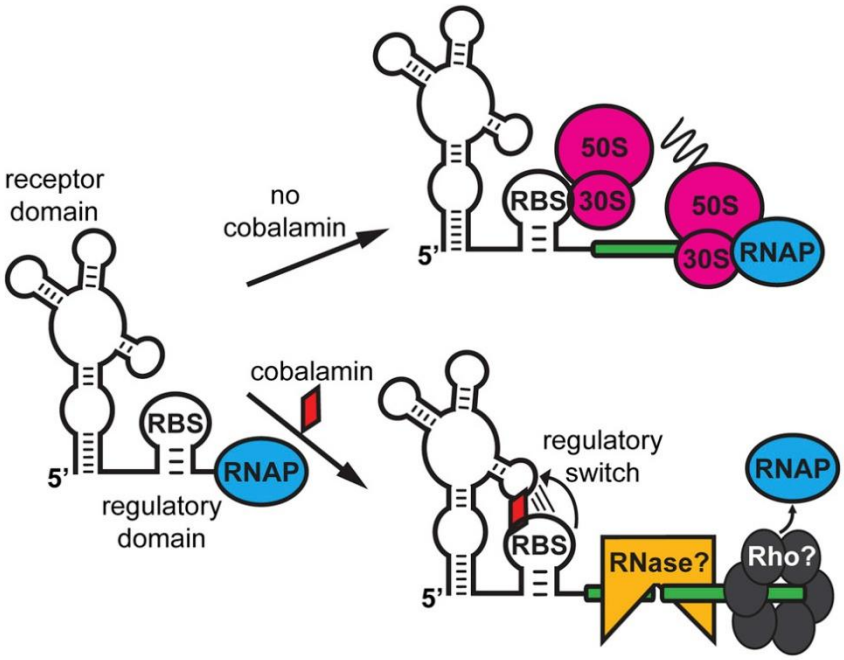

Figure 3 (Polaski, Jacob T., et al. 2016): general depiction of a cobalamin riboswitch mechanism. Upon ligand binding, regulatory elements are recruited to repress gene expression. Without cobalamin, the expression machinery is recruited normally.

molecules found within a cell to bind its ligand of interest. In fact, two classes of riboswitches have been described that discriminate between guanine and adenine despite an extremely high degree of homology in both their primary and secondary structures[11]. In essence, the receptor domain acts in the place of a sensory protein in the aspect that it can distinguish its sensor molecule with high specificity.

Upon ligand binding, the regulatory domain can “switch” between a genetic on state (ON state) which enables gene expression by preventing or occluding the formation of regulatory elements, or it can be in the genetic off state (OFF state) which represses gene expression forming an intrinsic terminator hairpin or a repression stem[5](Figure 3).

The mechanism by which the riboswitch directs regulation of gene expression has been a topic of study for the past few years. To give an example, the cobalamin riboswitch regulates gene expression by sequestering the ribosomal binding site (RBS) upon binding of its effector ligand, cobalamin[10].

Cobalamin Riboswitch

The cobalamin riboswitch represents a widely distributed class containing highly complex and independent riboswitches. Originally, all cobalamin riboswitches were proposed to bind adenosyl-cobalamin (AdoCbl), however, riboswitches capable of binding methyl-cobalamin and aqua-cobalamin were found[12]. This selectivity is primarily mediated by the interaction between a peripheral element of RNA that forms a T-loop module and a subset of nucleotides in the cobalamin-binding pocket[13]. Crystal structures of the env8 cobalamin riboswitch have been solved (Figures 5 and 6) and have revealed the core ligand binding pocket is defined by stands J3/4 and J6/3 of the four-way junction that is organized by a universally conserved T-loop mediated tertiary contact between L4 and L6 (Figure 5)[13].

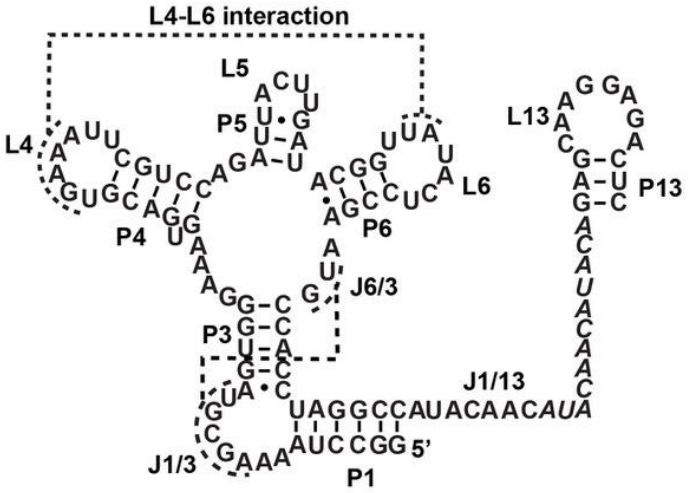

Figure 4: Secondary structure of the env8 Cobalamin riboswitch. Dashed lines depict nucleotides interacting to form the binding pocket for cobalamin. The 19 nucleotide polylinker is included in this motif which allows a greater range of fold repressions.

The regulatory mechanism of the cobalamin riboswitch is due to the interaction between the aptamer domain and the expression platform. Cobalamin itself, binds to the aptamer domain initially and forms an interaction with the expression platform containing the structural switch (Figure 4). Binding of cobalamin is primarily comprised of Van der Waals shape complementarity and a few direct hydrogen bonds[12]. The

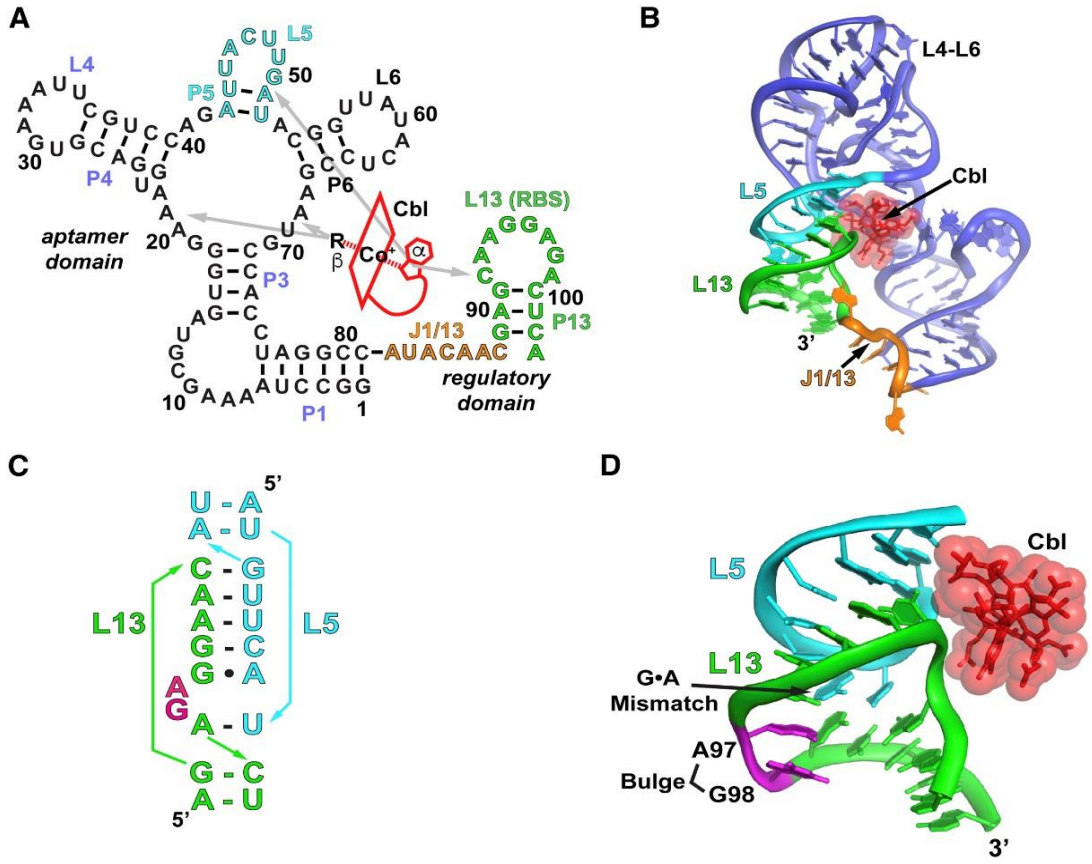

Figure 5 (Polaski, Jacob T., et al. 2016): (A) The secondary structure of the env8 cobalamin riboswitch with binding sites of CNCbl labelled with grey arrows. Colored T-Loops depict the motifs involved in kissing loop interaction. (B) Tertiary structure of the env8 cobalamin riboswitch with CNCbl bound to the construct. (C) Depiction of kissing loop motif with bulged nucleotides which destabilize the normally stable structure. (D) Zoomed in view of the kissing loop motif with CNCbl present to stabilize the structure. Bulged nucleotides are present which create the CNCbl dependence.

ligand is "sandwiched" between the P3-P6 coaxial stack and the third helix kissing-loop (KL) is created. The KL motif directly mediates regulation of the gene expression

machinery. Once cobalamin is bound to the riboswitch, the KL motif is formed between the two domains which occludes the RBS, thus repressing translation of the mRNA downstream of the riboswitch. This RNA module is formed through Watson-Crick base pairing of two terminal hairpin loops, between P5 of the receptor and P13 of the regulatory domain, to form a third helix (KL) that stacks between the helical stems of the two hairpins (Figure 5B)[10]. Kissing-loops are inherently stable structures widely used to promote tertiary architecture formation and RNA-RNA interactions[14]. A unique feature of the KL within the cobalamin riboswitch is the presence of non-Watson-Crick pairs and bulged nucleotides in the helix formed by pairing L5 and L13, thus creating an unstable KL[14].

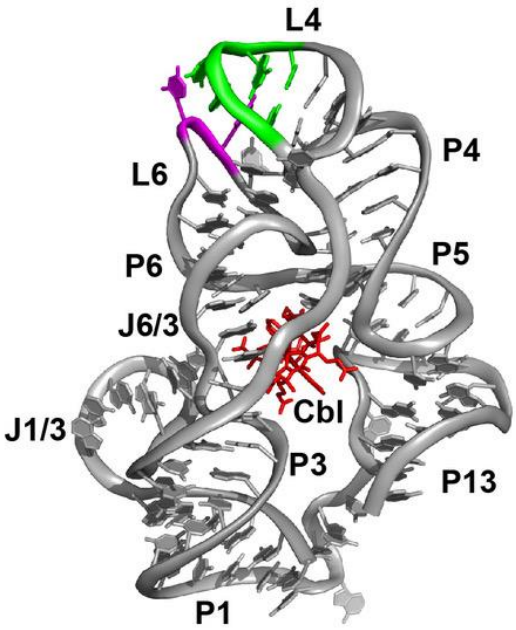

Figure 6: Tertiary structure of the cobalamin riboswitch. The riboswitch is in complex with CNCbl along with L4-L6 T-Loop interaction. The L4-L6 interaction supports the cobalamin binding pocket.

Destabilizing the KL interaction creates a dependence on cobalamin binding, as the ligand stabilizes the structure. To show bound cobalamin promotes the formation of the KL, its structure was probed as a function of Mg^{2+}, which promotes long-range tertiary interactions in RNA[12]. The results of the experiment concluded that when the riboswitch is under conditions of 15 mM Mg^{2+} the reactivity reaches sufficient results to conclude that KL formation has ensued. However, with cobalamin present the formation

of the KL occurs at physiological Mg^{2+} conditions (0.5 -1 mM) displaying the dependence upon cobalamin for KL formation[12]. Cobalamin stabilization of the KL module allows the riboswitch to attain its high degree of specificity along with the proper regulatory mechanism. Occlusion of the RBS is achieved quite simply through the KL motif sequestering it and bringing it into a conformation that is unable to bind the translation machinery.

The cobalamin riboswitch is dependent upon its tertiary structure and therefore must have RNA motifs that promote and stabilize it. T-Loops, through their ability to mediate long-range intermolecular and intramolecular interactions, provide this support.

T-Loops

T-Loops are defined as a 5-nucleotide motif found within the non-coding regions of RNA. These T-Loops provide two essential functions: they mediate long-range intermolecular interactions and they also facilitate intramolecular interactions between RNAs or a ligand of interest[22].

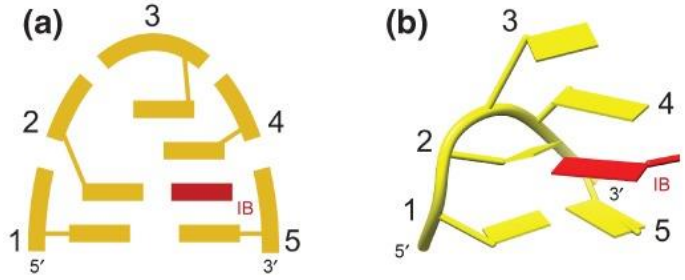

Figure 7 (Chan, et al. 2013): A diagram depicting the basic structure of the T-Loop motif. (a) a 2-dimensional representation showing the 5 nucleotide T-Loop including an intercalating base (IB). (b) a 3-dimensional representation of the T-Loop motif once again including an intercalating base to give an example of long-range interactions.

Within the context of a riboswitch, depending on the T-Loop, this RNA motif promotes the formation of the tertiary structure of the riboswitch. In addition to this functionality, the T-Loops within the riboswitch can play a role in ligand binding for riboswitch functionality. While the overall requirement of 5 nucleotides is conserved throughout all

T-Loops, the consensus sequence varies based upon the type of RNA that it is located within. Meaning, T-Loops are not sequence specific, but rather they are structure specific in accordance to the RNA structure required[22].

This leads us into the fundamental question of this research. What is the effect of variation upon a riboswitch T-Loop? Specifically, how are the intramolecular and intermolecular interactions of the L4 and L6 T-Loops affected when nucleotides within the motifs are randomized. Will be there a pattern of sequence conservation similar to the general riboswitch T-Loop consensus, or will it resemble another type of T-Loop structural motif?

Chapter 2: Methods

PCR Construction of env8 Cobalamin Riboswitch

The DNA library was constructed using two single-stranded Ultramers ® synthesized by Integrated DNA Technologies. These two single-stranded DNAs comprise the entire *env8* riboswitch, including the 7 randomized positions (16,384 variants). In addition to the 7 randomized positions, NsiI and HindIII restriction sites are included at the 5' and 3' ends of the construct. The two single-stranded ultramers were annealed and filled in by a single round of PCR using a Q5 DNA polymerase to obtain a dsDNA fragment to be cloned in later steps. An E.Z.N.A. PCR purification kit was used to isolate the DNA insert for later cloning. Following this procedure, the concentration (ng/ µL) of the DNA was measured by Thermo Scientific ™ NanoDrop ™ spectrophotometer.

DNA Cloning

After purification of the env8 cobalamin riboswitch library PCR insert, the DNA was digested with restriction enzymes HindIII-HF and NsiI-HF in a reaction containing 45 µL purified DNA product, 6 µL 10x CutSmart Buffer, 1 µL HindIII-HF enzyme, 1 µL NSI-HF enzyme, and 7 µL di-H_2O. The reaction mixture was then incubated for 1 hour in a 37 ^oC bath. Single digest controls consisted of 10 µL of DNA insert digested with HindIII only, 10 µL of DNA digested with NsiI only, and 5 µL undigested control. These were used to verify that the sample DNA has indeed been digested and successfully cut in the double digest by comparing lengths of DNA segments by gel electrophoresis (using 2% agarose). New reporter plasmid was cut with the same restriction enzymes to ensure complementary sticky ends for ligation. Restriction enzyme sites (NsiI and HindIII) within

the pRR5-mNeon reporter plasmid are located upstream of the mNeon gene. The pRR5-mNeon reporter plasmid digest procedure is as follows: 10 µL DNA insert (180 ng/ µL), 6 µL of 10x CutSmart Buffer, 1 µL of HindIII-HF, 1 µL of NSI-HF, 41 µL of di-H_2O, and 1 µL Alkaline Phosphatase, Calf Intestinal (CIP). The reaction mixture is allowed to incubate at 37 ^oC for 1 hour. CIP ensures that the 5' and 3' ends of the reporter plasmid do not re-ligate and close off the plasmid from ligation with the riboswitch insert. Following the 1-hour incubation, DNA insert and controls are run on a 2% agarose gel to ensure the restriction enzyme digest is completed, then it is purified using the E.Z.N.A. Gel Extraction kit to obtain a pure sample of DNA insert for following procedures.

Following the digestion protocol, the purified DNA insert is ligated into the pRR5-mNeon plasmid vector that has the appropriate HindIII and NsiI restriction sites. The reporter plasmid contains the fluorescent protein mNeon[16] in addition to an ampicillin resistance cassette (Amp^R). Procedure for the ligation is as follows: a reaction mixture of 2 µL 10x T4 Ligation Buffer, 1 µL T4 Ligase, 1 µL (180 ng/µL) pRR5-mNeon plasmid vector, 4 µL (20-40 ng/µL, purified concentration) DNA insert (diluted 1:4 from purified concentration), and 12 µL di-H_2O. The reaction mixture is then incubated at 16 ^oC for 24 hours.

The resulting DNA insert and pRR5-mNeon plasmid vector is then transformed into a chemically competent strain of *E.coli* BW25113 (ΔbtuR). Transformation is done by adding 3 µL of the ligation reaction to 100 µL of the ΔbtuR *E.coli*. The reaction mixture is allowed to rest on ice for 15 minutes, heat shocked (~40 ^oC) for 30 seconds, allowed to recover for 5 minutes with 900 µL of 2xYT media, and finally, the reaction mixture is incubated for 1 hour in a 37 ^oC water bath. After incubation, 100 µL aliquots of the

transformation reaction are plated onto 60 mm diameter plates filled with CSB media[17] containing 1.2% agar supplemented with cyanocobalamin (5 µM concentration) along with 100 µg/mL carbenicillin. Three controls are generated through transforming the ΔbtuR cells with the following: pRR5-mNeon plasmid containing the WT-env8 cobalamin riboswitch, GAAA control plasmid, or pBr322 plasmid. The transformed cells are allowed to incubate overnight (~19 hours) at 37 ^{o}C. To generate colony counting statistics, OpenCFU[18] software was used to count the number of colonies per plate once the colonies are grown.

Fluorescence Screening

A fluorescence screen was used to qualitatively isolate colonies with functioning riboswitches. Colonies are illuminated with a light source outfitted with 490 nm excitation and 510 nm emission filter to visualize the green fluorescence produced by the mNeon fluorescent protein. Colonies fit for isolation must exhibit a repression of fluorescence similar to wild type env8 riboswitch; upon meeting this requirement, the colony was picked using a sterile pipette tip and transferred to two separate grid plates. One grid plate will contain CSB medium containing 5 µM cyanocobalamin and 100 µg/mL carbenicillin while the other grid plate will only contain CSB media and antibiotic. In addition to a comparison between colonies containing the parental riboswitch, colonies are compared to control colonies containing the GAAA reporter plasmid. The GAAA plasmid contains the mNeon gene but with a defective env8 riboswitch included upstream of the gene, which will produce high fluorescence. Therefore, colonies exhibiting fluorescence that is less than this control are also isolated in this process so that low activity riboswitches are isolated

along with high activity. Once picking has been completed, the grid plates are incubated overnight (~19 hours) for colony growth at 37 ^{o}C.

Selection of Colonies from Grid Plates

Colonies from grid plates are selected in such a way that false positives are not included in downstream analysis. Fully functioning riboswitches will repress translation of

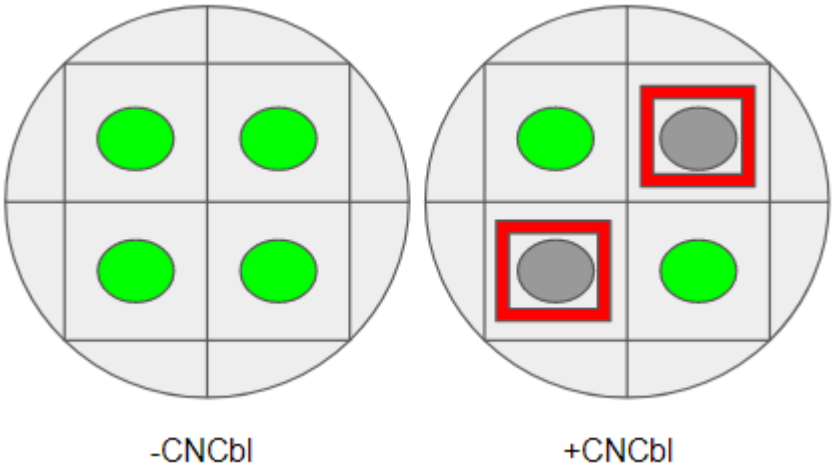

Figure 8: depiction of grid plates containing colonies with variant riboswitch ligated into pRR5-mNeon plasmid vector. Colonies squared are isolated during this phase of screening for fold repression testing due to the "switching" of the colonies in -CNCbl and +CNCbl environments

the mNeon fluorescent protein in the presence of cobalamin. However, the switch will also permit translation in the absence of the ligand. Thus, colonies with functioning riboswitches should express high levels of fluorescence on grid plates without cyanocobalamin and exhibit low levels of fluorescence on grid plates with cyanocobalamin (Figure 8)

Activity Assays

To determine the fold repression of an individual variant, colonies were picked from the grid plate without cyanocobalamin (via sterile pipette) and transferred to a liquid culture tube containing 3 mL of CSB medium supplemented with 100 µg/mL ampicillin. Then, 3 µL of each overnight culture were transferred to two separate tubes containing 3 mL of CSB medium with or without cyanocobalamin (5 µM) supplemented with 100 µg/mL of ampicillin. The culture tubes were grown for 5 hours or until they reach mid-log phase (cell density ranges of 0.5 to 0.6) at 37 ^{o}C in a roller drum. Once mid-log phase has been reached, 300 µL of each culture tube were transferred to a Costar ®96 well plate. The plate will also contain pRR5 plasmids with either the WT env8 riboswitch, pBr322, or GAAA controls along with a fluorescein standard (50 µL of 3 µM fluorescein and 250 µL di-H_2O). Controls are grown in the same manner as test cultures for consistency.

Cobalamin-dependent expression of the fluorescent protein was measured at an excitation wavelength of 490 nm and a 517 nm emission wavelength using a Tecan Infinite M200 ®PRO plate reader. Level of mNeon expression was determined in biological triplicate, which is then normalized to the cell density (OD_{600}). Normalized expression is then corrected by subtracting fluorescence observed in wells containing the pBR322 plasmid control to remove background fluorescence. Finally, fold repression is obtained by diving the activity seen in wells without cyanocobalamin present by wells with cyanocobalamin (i.e. unrepressed cultures divided by repressed cultures).

Mini-Prep DNA Sequences

5’ end of Ultramer	CACGACATGCATAAGGCTCGTATAATATATTCATATAATAATGGCCTAAAAGCGTAGT GGGAAAGTGACGNNNNNTTCGTCCAGATTACTTGATACGG
3’ end of Ultramer	ACAGGAAAGCTTGGCGTAATCATCAGCATGTTGAGTCTCCTTGCTCTGTATGTTGTA TGTTGTATGGCCTAGGTGGCATTCGGAGTANNACCGTATCAAGTAATCTGGACG
pRR5 Primer	GCGCTAGCCACAGCTAACAC

Figure 9: sequences of ultramers containing the env8 cobalamin riboswitch. Randomized positions are denoted by the letter “N”. pRR5 primer is included for sequencing the isolated plasmid.

Plasmids are isolated for sequencing after fold repression assays are completed to ensure that only functional riboswitches are recorded and accumulated together. To do this, cell cultures are grown in 5 mL of 2xYT media overnight (~16 hours) supplemented with 100 µg/mL ampicillin. These cultures are then spun down via centrifugation (3,000 rpm, 4 ^{o}C, 20 minutes) to pellet the cell mass and the supernatant is discarded. The plasmid is then isolated using an E.Z.N.A. Plasmid Mini-prep kit, once isolated a pRR5 sequence specific primer (Figure 9) is added to the isolated plasmid and it is sent in for sequencing.

Data Analysis

Organization of sequences within the FASTA format is set for highest to lowest fold repression to show differences in sequences based upon their functionalities. Once this format is achieved, it is entered into the Jalview program and variant positions are manually aligned. Riboswitch sequences are then separated into well-defined bins according to their functionality being high (>75% activity), moderate (<75% activity to >25%activity), or low (<25% activity). From the Jalview format, a WebLogo[20] cartoon is created to aid visualization of the consensus sequence. In addition to alignment according to magnitude of old repression, riboswitch sequences can be aligned according to

relatedness by using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) program in Jalview.

Results

Design of the Screen

To generate a functionally active library of L4 and L6 T-loop variants of the env8 cobalamin riboswitch, nucleotides 29-33 on L4 T-loop along with nucleotides 60 and 61 on the L6 T-loop were fully randomized to yield a total of 16,384 theoretically possible

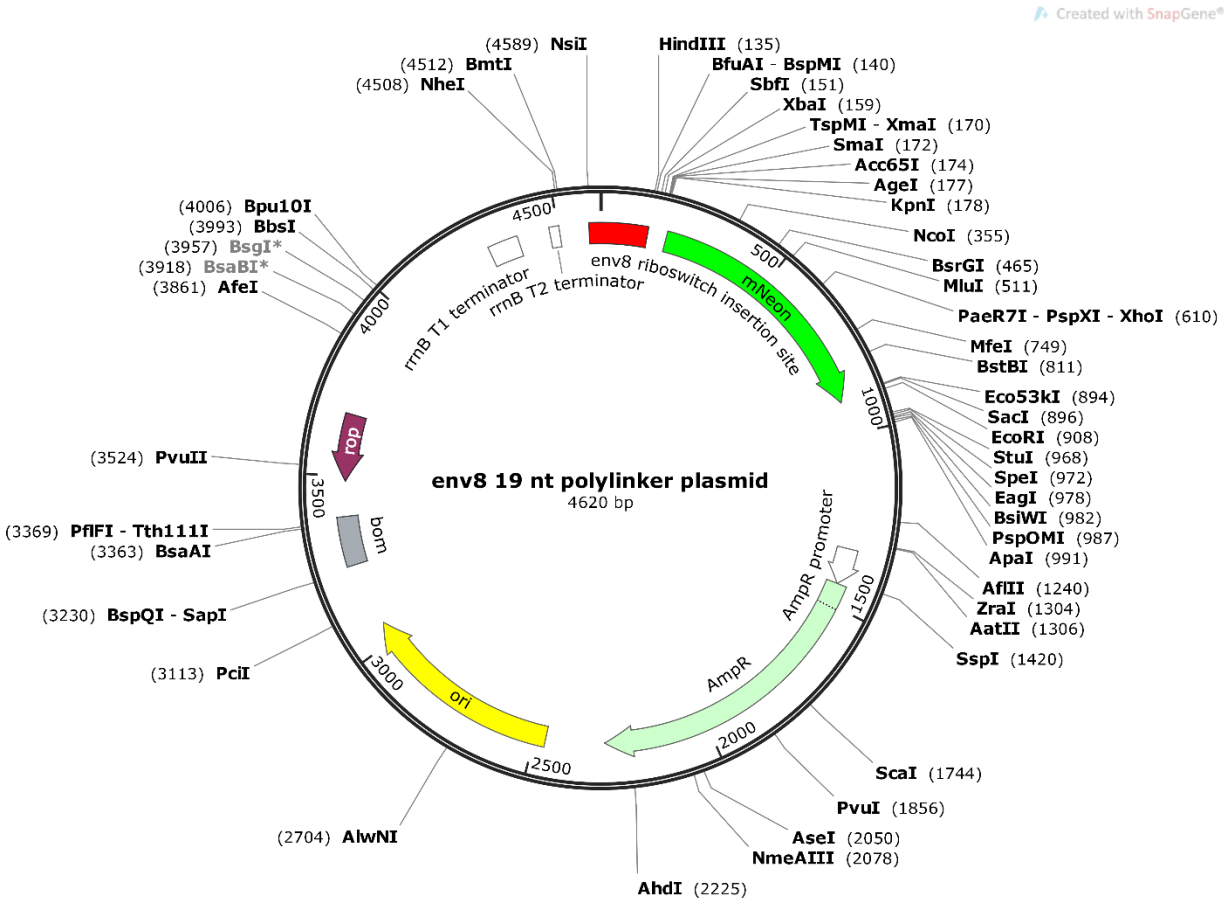

Figure 10 (generated through SnapGene): Plasmid map of the pRR5-mNeon vector displaying all restriction sites. mNeon is represented in green with the riboswitch insertion site within the leading region of the gene. Riboswitch insertion site is depicted in red to distinguish it from the mNeon gene

variants (Figure 11). The generated variant library was then cloned into a reporter plasmid that contained the gene for mNeon, a fluorescent protein that would be used for

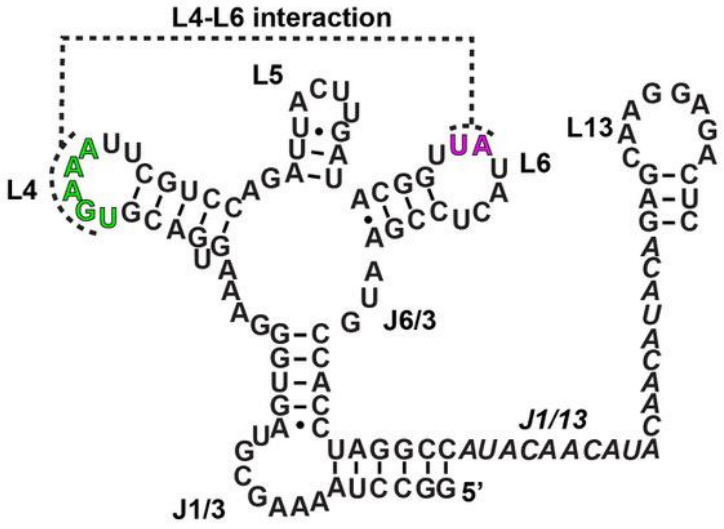

Figure 11: The env8 cobalamin riboswitch secondary structure. Locations of randomized positions are depicted by colored nucleotides. Green nucleotides are randomized positions within the L4 T-Loop and Purple nucleotides are within the L6 T-Loop.

qualitative analysis of functioning riboswitches (Figure 10). This reporter plasmid has been used in similar studies of env8 cobalamin riboswitches by Jacob Polaski (refer to reference 13). In addition to this randomization, the linker sequence connecting the regulatory and receptor domains was extended from 7 nucleotides to 19 nucleotides. The purpose of extending the linker sequence length is to increase the fold-repression range of the WT-*env8* riboswitch. This variant library was then cloned into *E. coli* BW25113 (ΔbtuR) strain; this strain of bacterium is unable to modify cobalamin upon import due to an induced defect within cob(I)yrinic acid a,c-diamide adenosyltransferase[19]. This defect ensures that the cyanocobalamin used in experimentation will not be converted into AdoCbl upon import, thus allowing the cobalamin riboswitch to bind CNCbl and elicit its regulatory response.

To screen for members of the randomized library that repress fluorescent protein expression in response to cobalamin, a bacterial culture transformed with the library

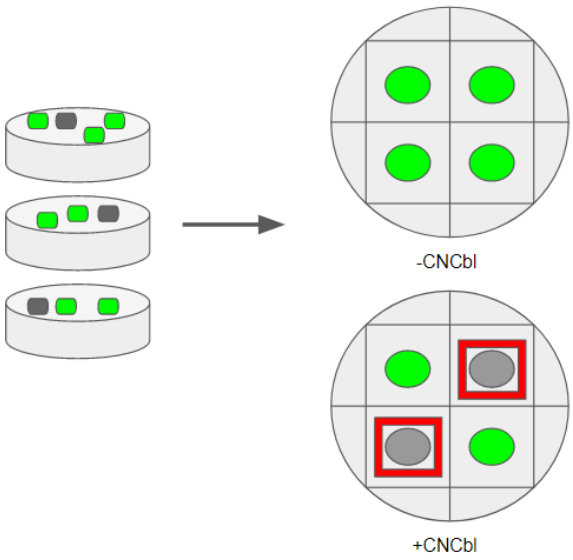

Figure 12: Depiction of screening process. Colonies displaying functional riboswitches from 1st round of screening (grey colonies) are transferred to grid plates shown to the right. Fully functional colonies will express the mNeon fluorescence protein in the absence of cobalamin but will repress expression in the presence of the effector ligand

was plated onto 60 mm diameter plates containing CSB agar containing 5 µM cyanocobalamin (CNCbl). To ensure sufficient colony separation while maximizing coverage, 50 µL of culture was be plated which yields approximately 100-200 colonies per plate. Plating onto CSB agar plates containing CNCbl was the first step in the screening process. The second step involved transferring colonies displaying probable functional riboswitches (dark colonies) to two separate grid plates (one containing CNCbl and one without). Functioning riboswitches were visually picked based upon repression of fluorescence that visually resembles colonies containing a pRR5-mNeon plasmid with the parental WT env8 riboswitch (Figure 12). Colonies from grid plates that displayed visual differences in fluorescence were used for quantitative measurement of their fold-repression values and sequencing.

In two rounds of screening, ~45,000 colonies were surveyed, representing an approximate 2.7-fold redundancy. This resulted in the likelihood that not all the sequences possible were seen in the two rounds of screening and therefore, further screening will need to be conducted to reach the desired ~99% of possible sequences being seen to conclude a consensus sequence attributed to functioning riboswitches. However, the sequences currently obtained are sufficient to yield an alignment that begins to show a pattern of sequences for functional riboswitches obtained from randomization of the L4 and L6 T-Loops.

Cobalamin Riboswitch Variants

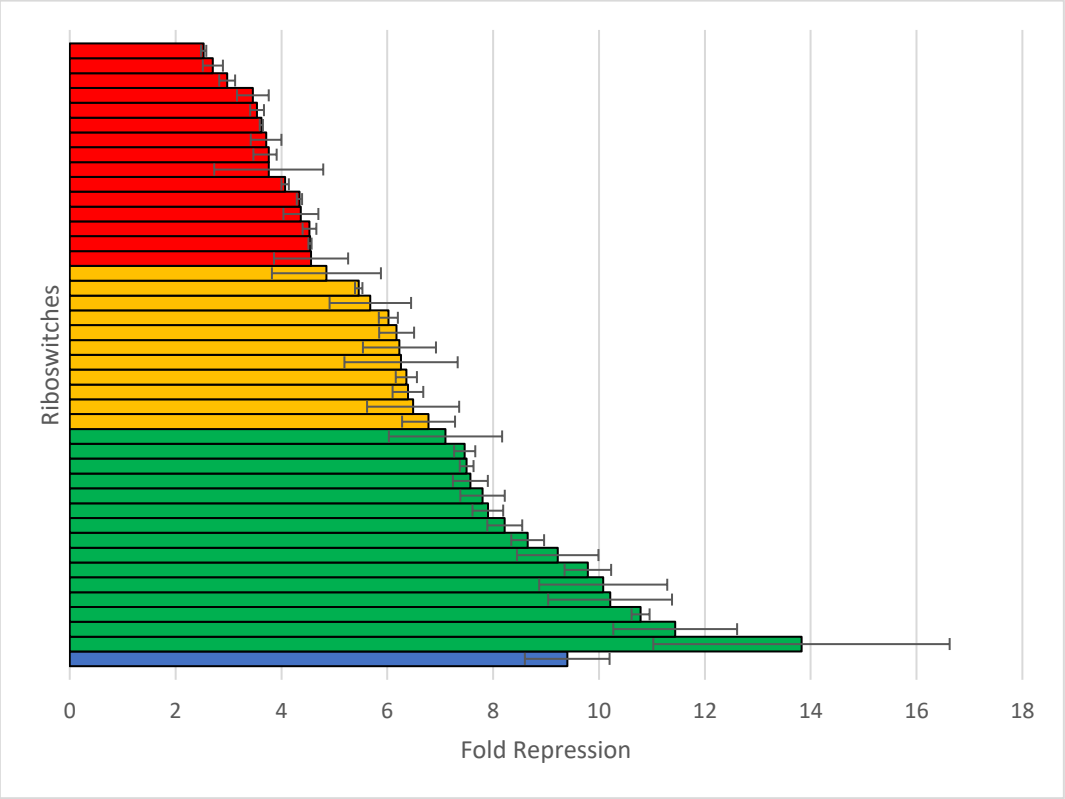

Figure 13. (A) Riboswitches depicting below 50% functionality relative to the WT env8 riboswitch are colored red (B) riboswitches depicting 50% to 75% functionality are colored yellow (C) riboswitches depicting above 75% functionality are colored green.

Variant riboswitch sequences were manually organized by fold repressions into a FASTA file that was inserted into Jalview to generate a sequence alignment. This was done to show sequence variation based upon fold repressions, thus generating an alignment that depicted patterns of variation that could display trends throughout various subsets of riboswitches.

Active cobalamin riboswitch variants that were isolated segregated into the following three classes: high, moderate, and low activity. High activity was classified as a riboswitch having >75% functionality when compared to the parent riboswitch (WT env8 cobalamin riboswitch), below 75% to 50% activity is deemed moderate, and below 50% is considered low activity. Analysis of the isolated sequences of the variant library was performed by organizing unique variant riboswitches by magnitude of fold repression in Jalview and comparing consensus sequences based on fold repression differences. Then a WebLogo was created for the ease of visualization of the consensus sequences.

Figure 14. WebLogo displaying the consensus sequence with all variant riboswitch sequences compiled together. This sequence can be a baseline depiction of what bases are required to maintain functionality of the cobalamin riboswitch.

High-Activity Riboswitches

This class of variant riboswitches is populated with sequences that exhibit repression activity that is either similar to the parental riboswitch (Figure 13) or surpasses the riboswitch functionality entirely. The consensus sequences of L4 and L6

Figure 15. Weblogo depicting the consensus sequence for all variant cobalamin riboswitches exhibiting fold repressions 75% or above that of the WT env8 parental riboswitch. Said riboswitches are termed "highly active"

T-loops show that the variant riboswitches have an invariant guanine at position 30, an invariant adenine at position 33, and an invariant adenine at position 61 (Figure 15). The nucleotides observed at these positions are consistent with the parental riboswitch sequence and therefore must be critical for efficient riboswitch functionality. Nucleotide position 29 was observed to prefer pyrimidines, specifically thymine. Positions between 31 and 32 on the L4 T-Loop show a preference for purine bases (82%) while position 60 on the L6 T-Loop displays a preference for pyrimidine bases (81%).

Moving into the interactions, positions 31-33 and 61 are involved in the base stacking between L4 and L6 T-Loops, position 61 being the intercalating base. All highly active riboswitches depict a conservation of purine bases at position 31 (82% purine), position 32 (100% purine), position 33 (100% purine), and position 61 (100% purine). This is likely because base stacking of purines is more stable than base stacking of

pyrimidines, therefore the preference for purines helps stabilize the L4 and L6 T-Loop interaction[21]. In addition, this class of riboswitches shows a preference for thymine and adenine at an intraloop Watson-Crick Hoogsteen base pairing of position 29 and 33. However, the presence of a C-A base pairing at this position coupled with high activity suggests that not all sequences have been seen in this subset of riboswitches. Observing shows the ribose-base-ribose interaction between positions 30, 32, and 60 displays preference for adenine at position 32 (64%) but, observing the overall bases seen at this position, only purine bases are included. This interaction forms an interaction between the L4 and L6 T-Loops, therefore the preference for adenine suggests that it forms a more stable complex between these positions. An invariant interaction seen in all sequences within this subset is the “sheared” G-A base pair located between positions 30 and 61. This interaction can be assumed to be imperative to the riboswitch function as these positions are consistently conserved.

Moderately-Active Riboswitches

This class of variant riboswitches is populated with sequences that exhibit repression activity between 75% and 50% of the WT env8 riboswitch fold repression (Figure 13). Similar to the parental riboswitch sequence, the invariance of nucleotide

Figure 16: WebLogo displaying the consensus sequence for riboswitches exhibiting a fold repression between 75% and 50% of the WT env8 cobalamin riboswitch parent.

positions 30 (guanine), 33 (adenine), and 61 (adenine) was seen. Again, this displays the importance of conserving these positions for riboswitch viability. At position 31, the consensus sequence for moderately active riboswitches shows a preference for guanine (56%) which is consistent with the overall preference for purines in the nucleotides involved in base stacking. This observation follows with the preference for purines in all positions involved in base stacking such as positions 32, 33, and 61. In addition to this, at position 60 on the L6 T-loop, the moderately active riboswitches show a preference for cytosine (64%) at this position whereas highly active riboswitches show little to no preference in terms of specific bases but do show a preference for pyrimidines.

Comparing the lowest value of repressions within the moderate subclass to the highest value fold repressions in the high activity subclass, we can begin to see a difference between the sequences that may provide insights into the effect of sequence variation. The sequences of both subsets conserved the bases at positions 30, 33, and 61. This shows the immense importance of these bases in the overall function of riboswitches. Discrepancies between the subsets appear in the levels of conservation at positions 29, 31, and 60. Moderate activity riboswitches display a greater level of variance at position 29 between cytosine and thymine, whereas highly active riboswitches show a preference for thymine. In addition to this, position 31 for moderate activity riboswitches shows a preference for guanine bases while highly active riboswitches prefer adenine. At position 60, highly active riboswitches show a preference for purines while moderate riboswitches show a greater level of cytosine compared to thymine.

Low-Activity Riboswitches

Figure 17: WebLogo displaying the consensus sequence for riboswitches exhibiting a fold repression below 50% and termed "low activity" sequences in relation to the WT env8 cobalamin riboswitch parent.

This class of variant riboswitches exhibits a fold repression that is comparable to less than 50% of the parental riboswitch and therefore has been deemed to have low activity (Figure 13, red region). Similarities between the low activity and the other subsets of riboswitches include the invariant position 30 guanine and the invariant position 61 adenine base. However, it was observed that at position 33, the normally invariant adenine was capable of being substituted with guanine. This variation was seen in a single sequence and therefore it can be assumed that not all functional sequences have been seen. In addition to this difference, position 32 has an inverse preference to the other two subset classes in that it shows a preference for guanine (55%) rather than adenine (45%). Although this inversion of preference does not affect the conservation of purine bases at positions included in base stacking between the L4 and L6 T-Loops, it does show that conservation patterns are not being followed at the low-activity riboswitches.

The pattern of non-specificity continues when observing position 60 (Figure 17). This position in low-activity riboswitches shows no specificity as displayed in the WebLogo. However, in the two upper classes of riboswitches a pyrimidine base was

favored at this position. Observing the interacting positions of 29 and 33, the base found at position 29 has an increased level of non-specificity and position 33 is still predominantly adenine. This observation may give insight into the reason for the decreased levels of repression found within this subset. The U-A Hoogsteen base pair characteristic of T-loops contributes highly to overall T-loop stability. Therefore, promoting variability at this position may destabilize the overall construct.

Conclusions on Variant Cobalamin Riboswitches

High-activity riboswitches and low-activity riboswitches seen here have a statistical significance in terms of difference in fold repression values regardless of the uncertainty in the measured value of fold repression (Figure 13). An important insight into the consensus sequence of cobalamin riboswitches as a whole is that positions 30, 33, and 61 must be conserved for any functionality to occur. In the low-activity subset, there is an increasing degree of variance in positions 29, 31, and 32. In the high-activity subset, there is a preference for purines at positions 31 and 32, while preferring thymine at position 29. The difference in positions 31 and 32 could provide an insight into the stability of the L4 and L6 interaction. As stated previously, purines provide a stronger base stacking interaction compared to pyrimidines, which would explain the preference displayed in high-activity riboswitches. However, low-activity riboswitches show a lower degree of preference for purines at these positions and therefore introduce a weaker interaction in the base stacking mechanism. With that being said, this displays that sequence variation at these positions does not disrupt riboswitch functionality entirely, but it does affect its effectiveness.

The main discrepancy between the riboswitches is the lack of conservation observed in low-activity riboswitches. Upon observation of the consensus sequences, low-activity riboswitches show complete non-specificity at positions 31 and 60, whereas in high-activity riboswitches there was a preference for purines at position 31 and pyrimidines at position 60. Position 31, being involved in base stacking, showed a preference for adenine in higher functioning riboswitches. However, this may describe a factor in the lower repression values seen in low activity riboswitches. A purine base in position 31 would create a less stable base stacking complex between the L4 and L6 T-Loop, therefore destabilizing the interaction and promoting the dissociation of the complex. Another discrepancy observed between the subsets can be seen between the base pairing between the position 32 base and position 60 ribose sugar. The inverse preference between the two subsets for position 32 (high activity preferring adenine and low preferring guanine) may give insight into how the L4 and L6 T-Loop interaction is stabilized. Meaning that guanine must not be able to stabilize the interaction as effectively as adenine by binding to the ribose sugars of positions 30 and 60.

Observing position 29, it can be inferred that the increasing preference for a thymine base as riboswitch functionality increases relates to the strength of the Hoogsteen base pairing with position 33. Highly functioning riboswitches strongly prefer the thymine base (82%) while the low-activity spectrum of sequences only contains the thymine base 55% of the time. With this being said, the riboswitch functionality is likely not attributed to a single entity being altered, but rather an interplay between multiple substitutions that provide a degree on instability. In other words, the functionality of a

cobalamin riboswitch depends upon the conglomeration of multiple substitutions or conservations.

Discussion

In the future, additional sequences must be obtained to ensure all possible variants have been observed. Currently, 39 sequences have been obtained ,which was sufficient enough to observe a trend within the consensus sequence that provides functioning riboswitches. If this trend continues and the invariant positions are maintained throughout the experiment, there should be a total of 64-72 potential variants and therefore additional screening is required to conclude a consensus sequence for the randomized L4-L6 interaction within the env8 cobalamin riboswitch.

The purpose of this study was to determine the effect of sequence variation within the L4-L6 interaction that make up a critical folding element of the aptamer domain in the *env8* cobalamin riboswitch. This was achieved using a cell-based screening of a library of over 16,000 RNA variants in which 7 nucleotides were fully randomized within the T-Loops of L4 and its two interacting nucleotide partners in L6 (Figure 11). Generating a complete randomization of these sites yields a diverse set of riboswitches which can be selected for function in an unbiased manner. Analysis of the sequences of the active variants of this library generated a consensus sequence that provides insight into the sequence requirements of the L4-L6 folding element of cobalamin riboswitch. The presence of high-activity and low-activity variants allowed for analysis into what nucleotides are needed for high functionality and what nucleotides are critical for the overall function in general. My approach to colony screening allowed for analysis of single variant riboswitches by studying the functionality of individual sequences and its comparison to the broader set of active sequences.

Comparison of the env8 L4 T-loop consensus to T-loops in other RNAs

Transfer RNA (47 T-Loops)	Position 1	Position 2	Position 3	Position 4	Position 5
	A = 2.1 %	A = 0 %	A = 2.1 %	A = 36.2 %	A = 97.9 %
	G = 0 %	G = 2.1%	G = 0 %	G = 63.8 %	G = 2.1 %
	C = 0 %	C = 0 %	C = 97.9 %	C = 0 %	C = 0 %
	U = 97.9 %	U =97.9 %	U = 0 %	U = 0 %	U = 0 %
	U	U	C	G	A

Riboswitches and Introns	Position 1	Position 2	Position 3	Position 4	Position 5
(11 T-Loops)	A = 9.1 %	A = 0 %	A = 72.7 %	A = 54.5 %	A = 81.8 %
	G = 0 %	G = 90.9 %	G = 0 %	G = 45.5 %	G = 9.1 %
	C = 9.1 %	C = 0 %	C = 27.3 %	C = 0 %	C = 0 %
	U = 81.8 %	U =9.1 %	U = 0 %	U = 0 %	U = 9.1 %
	U	G	A	A/G	A

L4 T-Loop Variant Riboswitch	Position 1	Position 2	Position 3	Position 4	Position 5
Consensus	A = 0 %	A = 0 %	A = 33 %	A = 57 %	A = 97 %
	G = 3 %	G = 100 %	G = 40 %	G = 43 %	G = 3 %
	C = 27 %	C = 0 %	C = 17 %	C = 0 %	C = 0 %
	U = 70 %	U =0 %	U = 10 %	U = 0 %	U = 0 %
	U	G	R	A/G	A

Figure 18: Tables above show the base compositions of various T-Loops along with their consensus sequence. The variant library depicts a strong resemblance to the riboswitch consensus sequence except for position 3, in which it only prefers purines. However, this is still consistent with typical riboswitch T-Loops as an adenine base is normally located at that position.

A comparison of the T-loop of L4 to the consensus sequences of other T-loop modules in various RNAs such as tRNA can give insight into the context dependency of T-Loops in terms of types of RNA and the functions it must perform. The T-loop itself is defined as a 5-nucleotide motif of a defined structure found within non-coding regions of RNA. These T-loops mediate long-range interactions as well as promote the tertiary structure of RNAs[22]. These motifs vary in base composition and consensus sequence when observing their occurrence in various forms of RNA. Comparing the L4 T-loop in the *env8* cobalamin riboswitch to the consensus sequence of tRNA T-loops, we observe that the L4 T-loop matches to consensus sequence for riboswitch T-loops to a relatively high degree (Figure 18), but there are differences. Differences between the typical riboswitch T-loop and the L4 T-loop are seen at position 2 and position 3 (Figure 18, middle table and bottom table). Position 2 within typical riboswitch T-loops is highly preferred to be guanine, however it can be substituted with uracil. The L4 T-loop has a strict requirement for guanine at this position, this is likely due to guanine having some effect on structure stability within the riboswitch. At position 3, typical riboswitch T-loops contain an adenine while the L4 T-loop is non-specific at this position. Within the context of the cobalamin riboswitch, this position is only involved in base stacking in the L4-L6 interaction and therefore may not have the utmost importance in terms of bases required.

When observing the T-loops found in tRNA and the L4 T-loop, differences arise within positions 2, 3, and 4. However, an important observation lies with position 1 and 2 within the three T-loop figures (Figure 18). These positions are conserved regardless of the RNA the motif is found within. Therefore, it can be assumed that this U-A base

pair is critical for the structure of T-loops. The difference occurring at position 2 is likely due to the structural requirements needed by the RNA module (Figure 18, top and bottom tables). Meaning, tRNA must have a degree of flexibility in translation due to the ratcheting mechanism of the ribosome. Therefore, the uracil base likely allows for greater flexibility compared to guanine which must better promote the structure needed of a T-loop within the riboswitch. Position 3 in the L4 T-loop is completely variable while within tRNA it is a highly conserved cytosine base. Position 4 in the L4 T-loop prefers purines while the tRNA T-loop conserves a guanine at this position. This difference in conservation patterns is once again, likely due to the structural requirements associated with either tRNA or riboswitches.

The consensus sequence for the L4 T-loop strongly resembles the overall riboswitch consensus sequence while at the same time it differs from the typical tRNA T-loop. This indicates that the sequence of nucleotides within a T-loop is dependent upon its function. For the *env8* cobalamin riboswitch, the sequence requirement is much different than the requirements for tRNA T-loops.

Comparison to J1/3 T-loop

A second T-loop in J1/3 and its interaction partner in J6/3 promotes the formation of the cobalamin binding pocket. The consensus sequence for the J1/3 and J6/3 T-Loop is significantly more forgiving in terms of variance with low activity riboswitches when compared to the L4 and L6 T-Loop interaction (Figure 19). While the L4 T-loop shows very little variance between the high activity and low activity riboswitches, there is a great amount of variance within the same subsets for the randomized J1/3 T-loop. This indicates that the L4-L6 T-loop mediated interaction is critical for riboswitch functionality while the J1/3-J6/3 interaction is less critical for riboswitch function. Put another way, these data at that indicate the T-loop mediated interaction of L4-L6 that mediates folding is essential for riboswitch function, while the J1/3 T-loop that facilitates ligand binding is not.

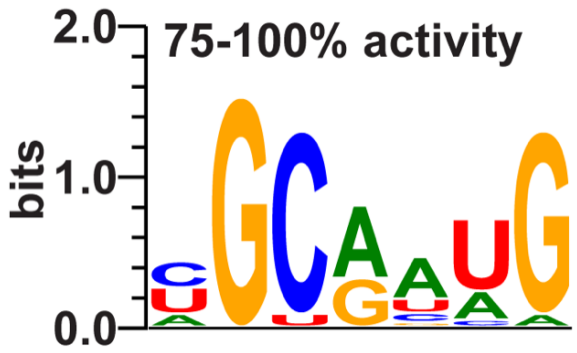

Figure 19 (Polaski, manuscript in preparation): A WebLogo of the consensus sequence for riboswitches depicting 75% or more functionality. This consensus is derived from a study that randomized nucleotides within the J1/3-J6/3 interaction. J1/3 is a T-loop while J6/3 is the interacting partner to promote the formation of the cobalamin binding pocket

This is an important insight into to how cobalamin riboswitches function and the importance of T-loop sequences within the aptamer domain. The fact that the consensus sequence for the L4-L6 interaction did not vary greatly between the low activity riboswitches and the high activity riboswitches shows the importance of the module within the aptamer domain. Couple this with the results of variation within the

J1/3-J6/3 interaction, it can be inferred that certain T-Loops of the aptamer domain play a less significant role than others.

Bibliography

1. Mandal, M., & Breaker, R. R. (2004). Gene regulation by riboswitches. *Nature Reviews: Molecular Cell Biology, 5*(6), 451-463. http://dx.doi.org.colorado.idm.oclc.org/10.1038/nrm1403

2. Ptashne, M. & Gann, A. Genes and Signals (Cold Spring Harbor Laboratory Press, New York, 2002

3. louin, S., Mulhbacher, J., Penedo, J. C. and Lafontaine, D. A. (2009), Riboswitches: Ancient and Promising Genetic Regulators. ChemBioChem, 10: 400–416. doi:10.1002/cbic.200800593

4. Sashital, Dipali G., and Samuel E. Butcher. “Flipping off the Riboswitch: RNA Structures That Control Gene Expression.” *ACS Chemical Biology*, vol. 1, no. 6, 2006, pp. 341–345., doi:10.1021/cb6002465. Accessed 23 Feb. 2018.

5. Gong, S., Wang, Y., Wang, Z., & Zhang, W. (2017). Computational methods for modeling aptamers and designing riboswitches. *International Journal of Molecular Sciences, 18*(11), 2442.

6. Barrick, Jeffrey E., and Ronald R. Breaker. "The Distributions, Mechanisms, and Structures of Metabolite-Binding Riboswitches." Genome Biology, vol. 8, no. 11, 2007, pp. R239-R239.

7. Winkler WC, Breaker RR: Regulation of bacterial gene expression by riboswitches. Annu Rev Microbiol. 2005, 59: 487-517.

8. Winkler WC: Metabolic monitoring by bacterial mRNAs. Arch Microbiol. 2005, 183: 151-159.

9. Tucker BJ, Breaker RR: Riboswitches as versatile gene control elements. Curr Opin Struct Biol. 2005, 15: 342-348.

10. Polaski, Jacob T., et al. "Mechanistic Insights into Cofactor-Dependent Coupling of RNA Folding and mRNA Transcription/Translation by a Cobalamin Riboswitch." *Cell Reports* 15.5 (2016): 1100-10. Web. 1 Mar. 2018

11. Noeske, Jonas, et al. "An Intermolecular Base Triple as the Basis of Ligand Specificity and Affinity in the Guanine- and Adenine-Sensing Riboswitch RNAs." Proceedings of the National Academy of Sciences of the United States of America, vol. 102, no. 5, 2005, pp. 1372-1377.

12. Johnson, James E. et al. "B_{12} Cofactors Directly Stabilize an mRNA Regulatory Switch." *Nature* 492.7427 (2012): 133–137. *PMC*. Web. 2 Mar. 2018.

13. Polaski, Jacob T., et al. "Cobalamin Riboswitches Exhibit a Broad Range of Ability to Discriminate between Methylcobalamin and Adenosylcobalamin." *The Journal of Biological Chemistry*, vol. 292, no. 28, 2017, pp. 11650-11658.

14. Butcher, SE, and AM Pyle. "The Molecular Interactions that Stabilize RNA Tertiary Structure: RNA Motifs, Patterns, and Networks." *Accounts of Chemical Research*, vol. 44, no. 12, 2011, pp. 1302-1311

15. Peselis, Alla, and Alexander Serganov. "Structural Insights into Ligand Binding and Gene Expression Control by an Adenosylcobalamin Riboswitch." Nature Structural & Molecular Biology 19.11 (2012): 1182-4. ProQuest. 4 Mar. 2018

16. Shaner, N. C., Lambert, G. G., Chammas, A., Ni, Y., Cranfill, P. J., Baird, M. A., . . . Wang, J. (2013). A bright monomeric green fluorescent protein derived from branchiostoma lanceolatum. *Nature Methods, 10*(5), 407-9.

17. Ceres, Pablo, et al. "Modularity of Select Riboswitch Expression Platform Enables Facile Engineering of Novel Genetic Regulatory Devices." ACS Synthetic Biology, 2013, pp. 130319131256007.

18. Geissmann, Quentin. "OpenCFU, a New Free and Open-Source Software to Count Cell Colonies and Other Circular Objects." *PLoS One*, vol. 8, no. 2, 2013;2012;, pp. e54072

19. Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., . . . Mori, H. (2006). Construction of escherichia coli K-12 in-frame, single-gene knockout mutants: The keio collection. *Molecular Systems Biology, 2*, 8-2006.0008.

20. Crooks, Gavin E., et al. "WebLogo: A Sequence Logo Generator." *Genome Research*, vol. 14, no. 6, 2004, pp. 1188-1190.

21. Sigel, Astrid, Bert P. Operschall, and Helmut Sigel. "Comparison of the π-Stacking Properties of Purine Versus Pyrimidine Residues. some Generalizations regarding Selectivity." JBIC Journal of Biological Inorganic Chemistry, vol. 19, no. 4, 2014, pp. 691-703.

22. Chan, Clarence W., Bhaskar Chetnani, and Alfonso Mondragón. "Structure and Function of the T-Loop Structural Motif in Noncoding RNAs: Structure and Function of the T-Loop Structural Motif." Wiley Interdisciplinary Reviews: RNA, vol. 4, no. 5, 2013, pp. 507-522.

23. Yanofsky, C. "Transcription Attenuation: Once Viewed as a Novel Regulatory Strategy."Journal of Bacteriology, vol. 182, no. 1, 2000, pp. 1-8.

24. McCown, Phillip J. Riboswitch Diversity and Distributions Suggest Ancient Origins, ProQuest Dissertations Publishing, 2014.

25. Polaski, Jacob T., et al., manuscript in preparation