

# **Contributions of Sequence and Structure to Ligand Selectivity in Class II Cobalamin Riboswitches**

**By: Samantha Webster**  
*Department of Chemistry and Biochemistry*  
*University of Colorado, Boulder*

**Defended March 29, 2016**

**Thesis Advisor:**

*Dr. Robert Batey, Department of Chemistry and Biochemistry*

**Thesis Committee:**

*Dr. Robert Batey, Department of Chemistry and Biochemistry*

*Dr. Joseph Falke, Department of Chemistry and Biochemistry*

*Dr. Thomas Perkins, Department of Molecular, Cellular, and Developmental Biology*

# Table of Contents

|   |    |
|---|----|
| <b>Abstract</b> .....   | 3  |
| <b>1. Introduction</b> .....  | 4  |
| 1.1 <i>Gene Regulation by Non-Coding RNAs</i> .....                   | 4  |
| 1.2 <i>Riboswitches</i> .....   | 5  |
| 1.3 <i>Cobalamin Riboswitches</i> .....                               | 7  |
| 1.4 <i>A Range of Binding Affinities in Cbl-II Riboswitches</i> ..... | 12 |
| <b>2. Methods</b> .....   | 16 |
| 2.1 <i>Construction of DNA Templates</i> .....                        | 16 |
| 2.2 <i>Transcription of DNA into RNA</i> .....                        | 16 |
| 2.3 <i>Purification of RNA</i> .....                                  | 17 |
| 2.4 <i>Isothermal Titration Calorimetry</i> .....                     | 18 |
| <b>3. Results and Discussion</b> .....                                | 21 |
| 3.1 <i>Mutant Design</i> .....  | 21 |
| 3.2 <i>Affinities of env50 Mutants for AdoCbl and MeCbl</i> .....     | 23 |
| 3.3 <i>Effect of env50 Mutations on Ligand Selectivity</i> .....      | 23 |
| 3.4 <i>Current Experiments with env8 Mutants</i> .....                | 28 |
| 3.5 <i>The Broader Picture of General Ligand Binding</i> .....        | 32 |
| <b>References</b> .....   | 34 |
| <b>Acknowledgements</b> .....   | 38 |

## Abstract

Riboswitches are gene regulatory elements found in the 5'-untranslated region of many bacterial genes. The direct binding of small molecule ligands induces conformational changes in the RNA that influence downstream expression machinery and determine the transcription or translation of an attached gene. Cobalamin riboswitches, which bind coenzyme-B<sub>12</sub> and other cobalamin variants, are divided into two distinct classes. Cbl-I riboswitches contain a large peripheral element that limits riboswitch binding to only coenzyme-B<sub>12</sub> (adenosylcobalamin, or AdoCbl). Cbl-II riboswitches do not contain the same peripheral element, and were originally thought to only bind smaller species of cobalamin (including methylcobalamin, or MeCbl). However, a sample of Cbl-II variants reveals a range of riboswitch binding affinities to both AdoCbl and MeCbl, with some RNAs binding both species of cobalamin with equal affinity. In order to determine the structural components of the Cbl-II riboswitches that are important in establishing ligand selectivity, a series of RNA mutants were designed to convert an AdoCbl-selective riboswitch (WT*env50*) to MeCbl-selectivity. Sequences from the J6/3 and J3/4 binding core, the J1/3 peripheral element, and the P4/P6 interaction were targeted, using a MeCbl-specific riboswitch (WT*env8*) as a guide. Isothermal titration calorimetry (ITC) results indicate that the J1/3 interaction with J6/3 is critical in conferring ligand selectivity in Cbl-II riboswitches. Ongoing experimentation on WT*env8* scaffold mutants with WT*env50* directed mutations might yield further insight into Cbl-II riboswitch ligand selectivity and into broader RNA strategies for general ligand binding.

## 1. Introduction

### 1.1 Gene Regulation by Non-Coding RNAs

One of the biggest surprises that has come from the human genome project is the discovery that most of the genome is non-coding. While more than 85% of the human genome is actively transcribed, only about 3% actually codes for proteins (1). A big question, then, is what is the point of all this extra transcription, and what are all these RNAs doing?

Several decades of study have gone into characterizing a host of non-coding RNA classes, which have been found to be intimately involved in the regulation and maintenance of the genome (2). While RNA is the physical intermediary between the coding genome and functional protein, it is also involved in genome upkeep and patterns of gene expression. Non-coding RNAs act at virtually every level of the central dogma and beyond, and are involved in everything from broad scale gene packaging to ribosome composition (2).

Direct control of gene expression by non-coding RNAs can be in *trans*, as with microRNAs, which bind other RNA molecules in order to silence them (3, 4), or in *cis*, with non-coding structural elements of an mRNA helping to determine whether or not the attached coding gene will be translated into protein (2, 5). Some of these structural elements were shown to bind attenuating proteins, which were in turn sensitive to intracellular metabolite concentration, in order to control gene expression, as with the tryptophan RNA-binding attenuation protein (TRAP) and the trp operon (5). Later, leading sequences of mRNAs were shown to bind small molecule ligands and direct

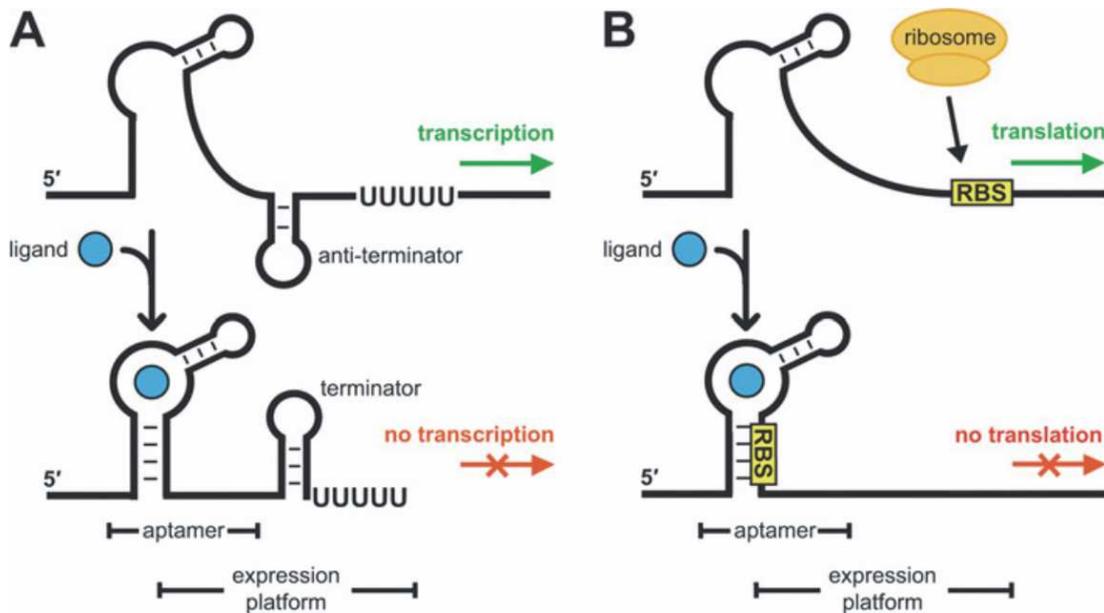
bacterial gene expression without the need for a mediating protein. These metabolite-binding mRNA elements were collectively termed riboswitches (6).

### *1.2 Riboswitches*

Riboswitches are a near-ubiquitous class of non-coding bacterial RNA that directly bind a range of small molecule ligands, including amino acids, coenzymes, and other fundamental metabolites (7,8,9). They are found in every major clade of bacteria, and allow cells to sense and respond to small molecule concentrations in their environment, participating in the regulation of genes necessary for cellular homeostasis, metabolism, and even virulence (10). Riboswitches are nearly always found in the 5' untranslated region of mRNAs, and consist of two functional domains: a receptor domain that can recognize and bind a small molecule ligand, and an expression platform, which responds to ligand binding by directing downstream expression machinery (11).

Riboswitches can control gene expression on both a transcriptional and translational level (9). In transcriptional control, the aptamer domain can assume one of two conformations as the riboswitch is transcribed, depending on the presence of its cognate ligand (12). In an "OFF" switch (shown in **Figure 1A**), the binding of a ligand stabilizes a terminator stem loop, which stalls the RNA polymerase and halts transcription. The absence of the ligand allows an anti-terminator loop to form, and transcription of the attached gene is allowed to continue (12). "ON" switches have also been observed, where ligand binding instead stabilizes the anti-terminator loop and allows transcriptional read-through (13).

In translational riboswitches, the expression platform contains the ribosome binding site (RBS), which can be sequestered by the aptamer domain upon ligand binding (as in an “OFF” switch, **Figure 1B**). While the RBS interacts with the aptamer domain, the ribosome cannot recognize or bind the mRNA, and protein translation does not occur (14).



**Figure 1: Transcriptional and translational attenuation mechanisms.** A) Transcriptional “OFF” switch. In the presence of ligand, a conformational change in the aptamer domain stabilizes a terminator loop in the expression platform, stalling RNA polymerase and inducing dissociation of the mRNA. B) Translational “OFF” switch. In the presence of ligand, a conformational change in the aptamer domain sequesters the ribosome binding site and does not allow for translation of the mRNA. Figure taken from reference 15.

Most riboswitches are both sensitive and highly specific for the small metabolites they can recognize and bind, discriminating between their cognate ligand and its chemical homologues at levels rival to proteins, even though RNAs have a far less diverse array of sequential building blocks to choose from (16-18). In the case of the lysine riboswitch, lysine is bound with at least a 10,000-fold higher affinity than other amino acids, including ornithine, which differs structurally from lysine by a single

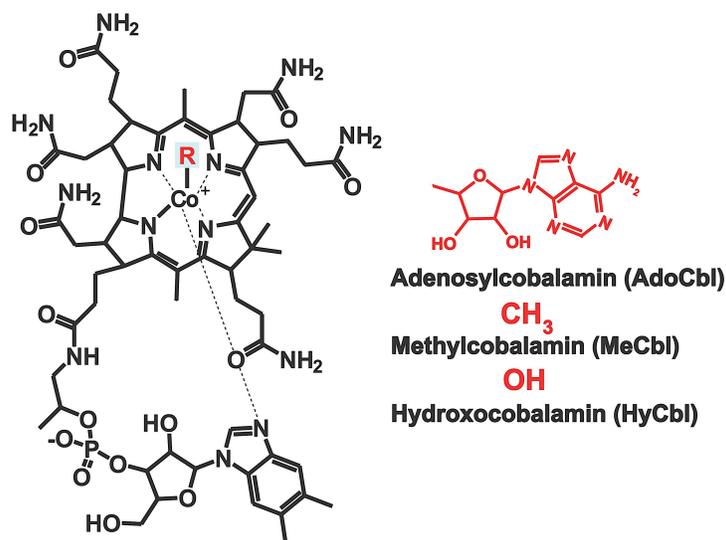
methylene group (19, 20). This discrimination is in spite of significant recognition of the main chain atoms of the amino acids (21, 22).

### 1.3 Cobalamin Riboswitches

The first validated riboswitch was found in the mRNA leader region of the cobalamin transport gene *btuB* in *E. coli* (23, 24). After it was observed that *btuB* expression was repressed in cells grown in the presence of cobalamins (24, 25), a “B12-box” in the mRNA was discovered to bind cobalamin directly, without the assistance of protein (26). Phylogenetic analysis showed widespread conservation of the sequence in prokaryotes (27, 28), and coenzyme B<sub>12</sub>-dependent riboswitches were found to control the expression of multiple genes, including *btuB* in *E. coli* and *S. typhimurium*, and genes of the cobalamin biosynthesis (*cob*) operon in *S. typhimurium* (23).

Coenzyme B<sub>12</sub>, or adenosylcobalamin, is a complicated small molecule cofactor that is involved in several crucial metabolic pathways, including methionine synthesis, the methylmalonyl-CoA mutase reaction, and the ribonucleotide reductase reaction (28). The structure of cobalamin consists of a corrin ring that coordinates a cobalt ion at its center; the fifth group coordinating the cobalt is a dimethylbenzimidazole (DMB) group extending from the corrin ring, while the sixth group (also known as the beta-axial moiety) varies among species of cobalamin. Three species of cobalamin exist in biological systems (**Figure 2**) (29). Adenosylcobalamin (AdoCbl) has a 5'-deoxyadenosyl (5' dAdo) moiety attached at the beta-axial position. The instability of the carbon-cobalt bond in AdoCbl allows rapid photolysis to hydroxocobalamin (HyCbl), which coordinates a hydroxyl group at the beta-axial position (30, 31). The homolytic

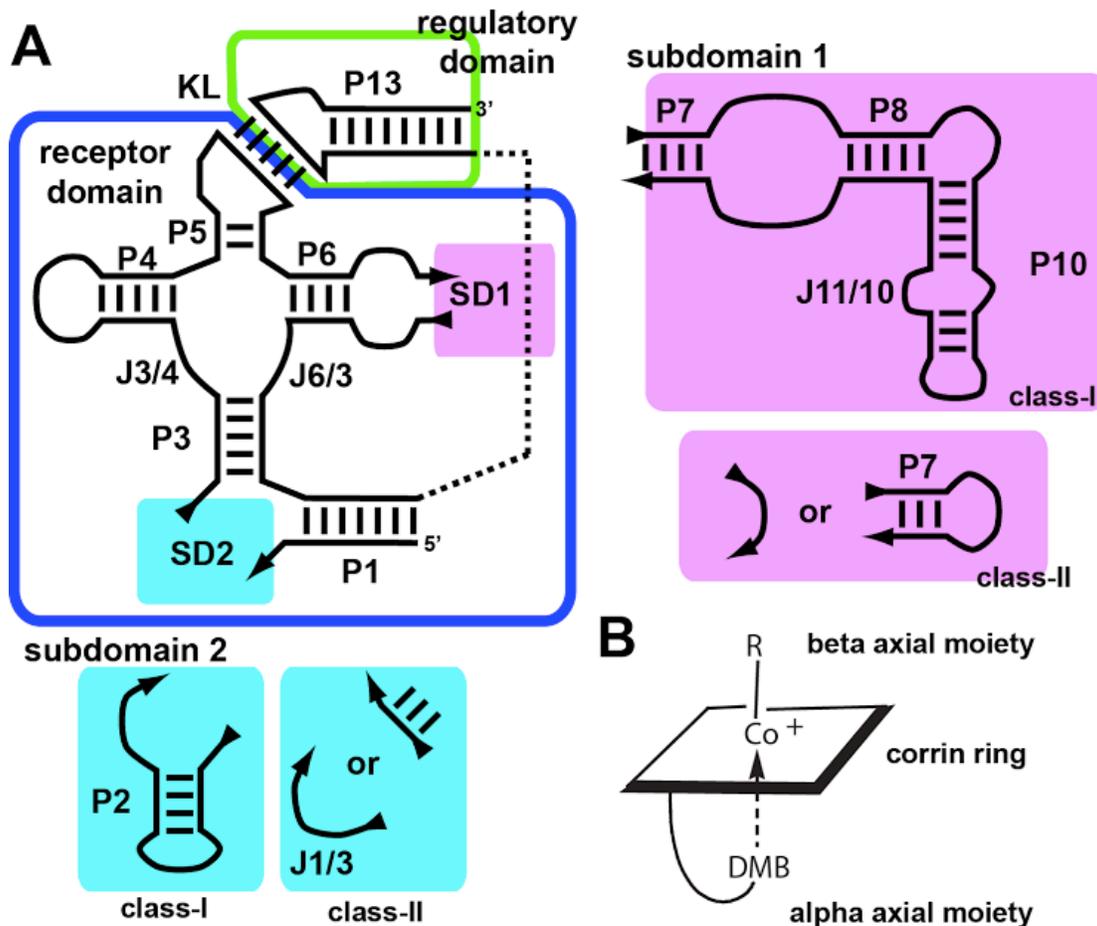
cleavage of AdoCbl yields a 5'-deoxyadenosyl radical, which plays an essential part in Coenzyme B<sub>12</sub>-catalyzed reactions (28). The third species is methylcobalamin (MeCbl), which is similar in size to HyCbl and coordinates a methyl group at the beta-axial position (29).



**Figure 2: Biologically relevant cobalamin species.** The ‘R’ group represents an interchangeable beta-axial moiety. Adenosylcobalamin (AdoCbl) coordinates a 5'-deoxyadenosyl group, and can rapidly photolyse to hydroxocobalamin (HyCbl) which coordinates a hydroxyl group. Methylcobalamin (MeCbl) coordinates a methyl group and is similar in size to HyCbl. Figure taken from reference 32.

All members of the cobalamin riboswitch clan contain a central four-way junction motif (**Figure 3A**, outlined in blue) in their receptor domains; this motif contains a number of highly conserved nucleotides, especially in the joining regions J3/4 and J6/3 (32, 33). Crystal structures of three cobalamin riboswitch receptor domains have shown that this region is critical to the interaction of the riboswitch with the beta-axial moiety of cobalamin (**Figure 3B**), as well as parts of the corrin ring face that are oriented in the same direction (30, 31). The DMB moiety and the opposite face of the corrin ring interact with a kissing loop (KL) between stem loop (SL) 5 and SL13 of the regulatory

domain, which typically contains the RBS. This interaction provides the basis for ligand-dependent translational control (32).

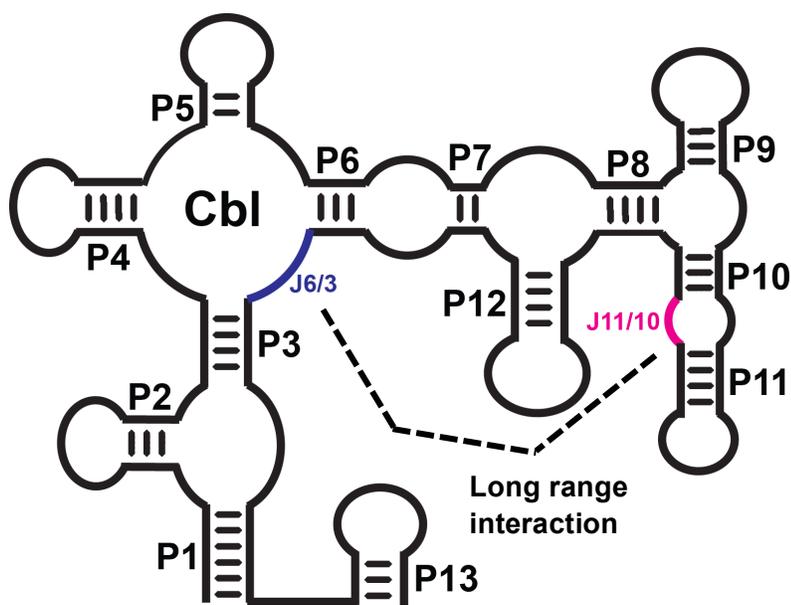


**Figure 3: Conserved secondary structure of cobalamin clan riboswitches.** A) The central four-way junction motif (outlined in blue) is conserved in all cobalamin riboswitches. It contains several highly conserved nucleotides, particularly in the J3/4 and J6/3 regions, and is important in the interaction with the beta-axial moiety and the top face of the corrin ring. The DMB moiety and the opposite face of the corrin ring interact with the KL between P5 of the central junction and P13 of the regulatory domain (outlined in green), which gives rise to ligand-dependent regulatory control. Subdomain 1 (pink) and subdomain 2 (blue) vary between the two classes of cobalamin riboswitches. B) A schematic of the cobalamin molecule, including the beta-axial moiety (top), the corrin ring, and the DMB alpha axial moiety (bottom). Figure taken from Rob Batey.

Based on phylogenetic and consensus sequence analysis, two classes of cobalamin riboswitches have been identified. 90% of the cobalamin riboswitch clan falls into the Class I (Cbl-I) category, which is distinguished by the presence of a large

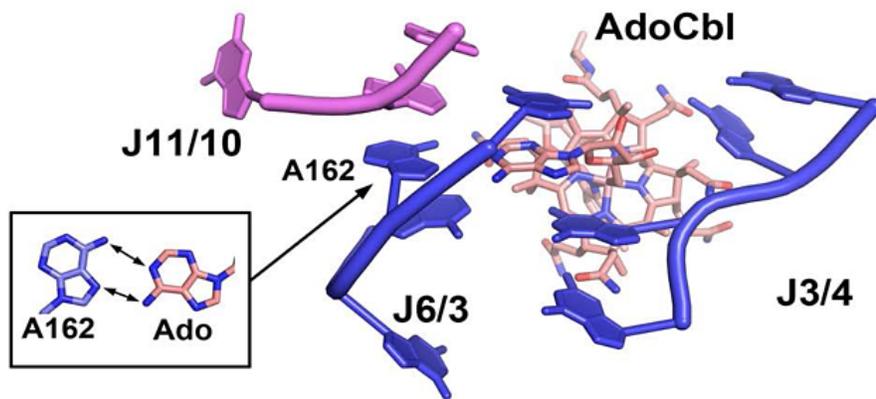
subdomain 1 (SD1) (**Figure 3**, pink) that consists of paired regions P8-P11 (27). This extension contains a purine-rich internal loop (IL) J11/10 (**Figure 4**, pink) that interacts with J6/3 of the central four-way junction (**Figure 4**, blue). In AdoCbl-binding riboswitches, J3/4 and J6/3 directly interact with the 5'-dAdo moiety via the 2' and 3' hydroxyl groups of the 5'-deoxyribose sugar, and via base-base interaction between the adenosyl group and A162 in J6/3 (**Figure 5**). The long-range interaction between J11/10 and J6/3 is thought to facilitate this contact with the 5'-dAdo moiety, as J11/10 helps to pull away A162 of J6/3 away from the binding core, and allow it to accommodate the bulky beta-axial group. Since the bases in J11/10 that interact with J6/3 are invariant across Cbl-I riboswitches, it's likely that all members of this group have high specificity for AdoCbl (32, 33). This is consistent with available biochemical

data (23, 24, 34, 35).



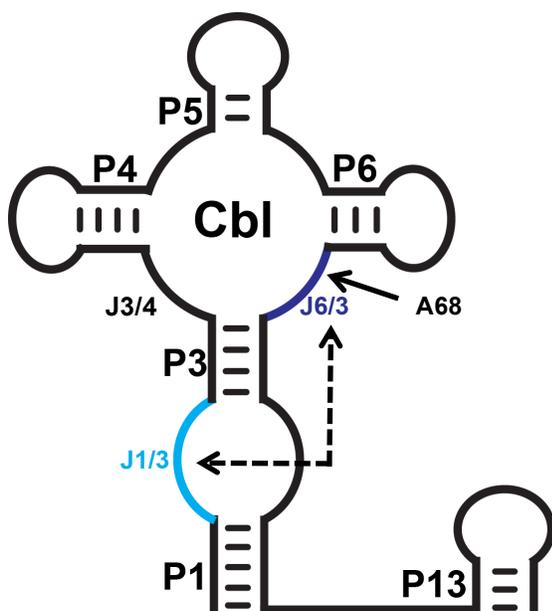
**Figure 4: General secondary structure of a Class I cobalamin riboswitches.** Cbl-I riboswitches are characterized by a large SD1, which consists of paired regions P8-P11. A conserved purine-rich IL J11/10 (pink) interacts with central junction J6/3 (blue) and facilitates the binding core's interaction with the 5'-dAdo moiety. Figure taken from reference 32.

The second class of cobalamin riboswitches (Cbl-II) is a more varied group, and includes members of "AdoCbl-variant" (from a Pacific Ocean metagenome) (36) and *Enterococcus* variant riboswitches (37). The class is defined entirely by the lack of the P8-P11 peripheral extension (**Figure 6**).



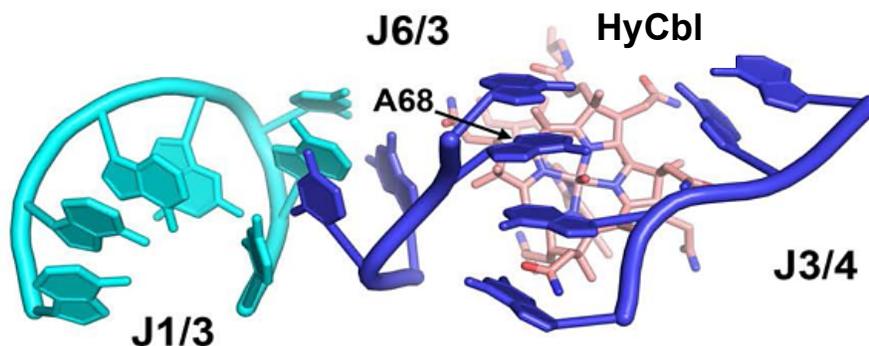
**Figure 5: Crystal structure of binding pocket of Class I riboswitch with AdoCbl.** J3/4 and J6/3 directly interact with the 5'-dAdo moiety via the 2' and 3' hydroxyl groups of the 5'-deoxyribose sugar, and via a base-base interaction between the adenosyl group and A162 in J6/3. A long-range interaction between J6/3 and J11/10 (pink) positions J6/3 to allow its contact with AdoCbl. Figure from reference 32.

The subdomain 2 (**Figure 3**, blue) also diverges from the Cbl-I riboswitches, though it varies within the Cbl-II group (27). The absence of the large SD1 peripheral extension, which is shown to help confer AdoCbl specificity in Cbl-I riboswitches, seems to suggest a higher affinity of Cbl-II riboswitches for the smaller species of cobalamin. In MeCbl-binding riboswitches, J3/4 and J6/3 interact through a purine-stacking interaction, which is thought to sterically occlude the 5'-dAdo moiety of AdoCbl (**Figure 7**). In the case of WT*env8* it has been shown that J6/3 is buttressed by peripheral element J1/3, and that



this may contribute to the occlusion of the binding pocket (32). However, J1/3 is highly variable within the Cbl-II group, and is even absent in some variants (36).

**Figure 6: General secondary structure of a Class II cobalamin riboswitch.** Peripheral extension P8-P11 is absent in Cbl-II riboswitches. J1/3 varies within the Cbl-II group. Dotted lines show interaction between J1/3 and J6/3 in WT*env8*. Figure taken from reference 32.



**Figure 7. Crystal structure of Class II *env8* binding pocket with HyCbl.** J3/4 remains in the same orientation as in AdoCbl-binding riboswitches, but J6/3 has shifted so it interacts with J3/4 via purine-stacking, with A68 packing against the beta-axial face of the corrin ring and sterically occluding the 5'-dAdo moiety. In this riboswitch, J6/3 is buttressed by peripheral element J1/3. Figure taken from reference 32.

#### 1.4 A Range of Binding Affinities in Class II Cobalamin Riboswitches

A sample of Cbl-II riboswitches shows that there is actually a wide range of cobalamin affinities within the group (**Table 1**). Some variants, like a Cbl-II riboswitch that controls the *eutTG* operon in *Enterococcus faecalis* (37, 38), show strong AdoCbl affinity and little to no binding to MeCbl or other cobalamins with small beta-axial moieties. Others, like the *env4* and *env8* riboswitches (from a Pacific Ocean metagenome) (36), are highly specific for MeCbl, and show no detectable binding to AdoCbl (32). There are also several variants that do not show strong specificity for either MeCbl or AdoCbl: *env50* is more selective for AdoCbl by about 100-fold, but still binds both compounds; *env62* cannot distinguish between the two species of cobalamin.

Most of these wild type riboswitches bind to at least one species of cobalamin at nanomolar concentrations. Environmental cobalamin concentrations range from 2-4 pM in open water, and cells are typically able to bioconcentrate cobalamin upwards of a thousand-fold from environmental levels (47). The intracellular cobalamin concentration

of *E. coli* has been estimated at upwards of 300 nM (48), though intracellular cobalamin concentrations likely vary widely between bacterial species, and may even vary between cellular compartments (47). It is likely that the dissociation constant reported for each wild type riboswitch indicates biologically relevant cobalamin concentrations for the species in question.

| riboswitch               | $K_D$ , AdoCbl<br>(nM) | $K_D$ , MeCbl<br>(nM) | $K_{rel}$ <sup>a</sup> |
|--------------------------|------------------------|-----------------------|------------------------|
| <i>env4</i> <sup>b</sup> | >50000                 | 19 ± 4                | >2600                  |
| <i>env8</i> <sup>b</sup> | >50000                 | 7.5 ± 4               | >6600                  |
| <i>env47</i>             | 950 ± 160              | 2300 <sup>c</sup>     | 0.41                   |
| <i>env50</i>             | 2.0 ± 0.58             | 190 ± 23              | 0.011                  |
| <i>env62</i>             | 105 ± 18               | 56 ± 21               | 1.9                    |
| <i>A. marina hupE</i>    | >50000                 | 2500 <sup>c</sup>     | >20                    |
| <i>S. elongates hupE</i> | 70 <sup>c</sup>        | 2.5 <sup>c</sup>      | 28                     |
| <i>T. elongates cbiX</i> | 11000 <sup>c</sup>     | 2300 <sup>c</sup>     | 4.8                    |
| <i>E. facealis eutG</i>  | 150 ± 30               | >50000                | <0.0030                |

**Table 1: Affinities of Cbl-II variants for AdoCbl and MeCbl.**

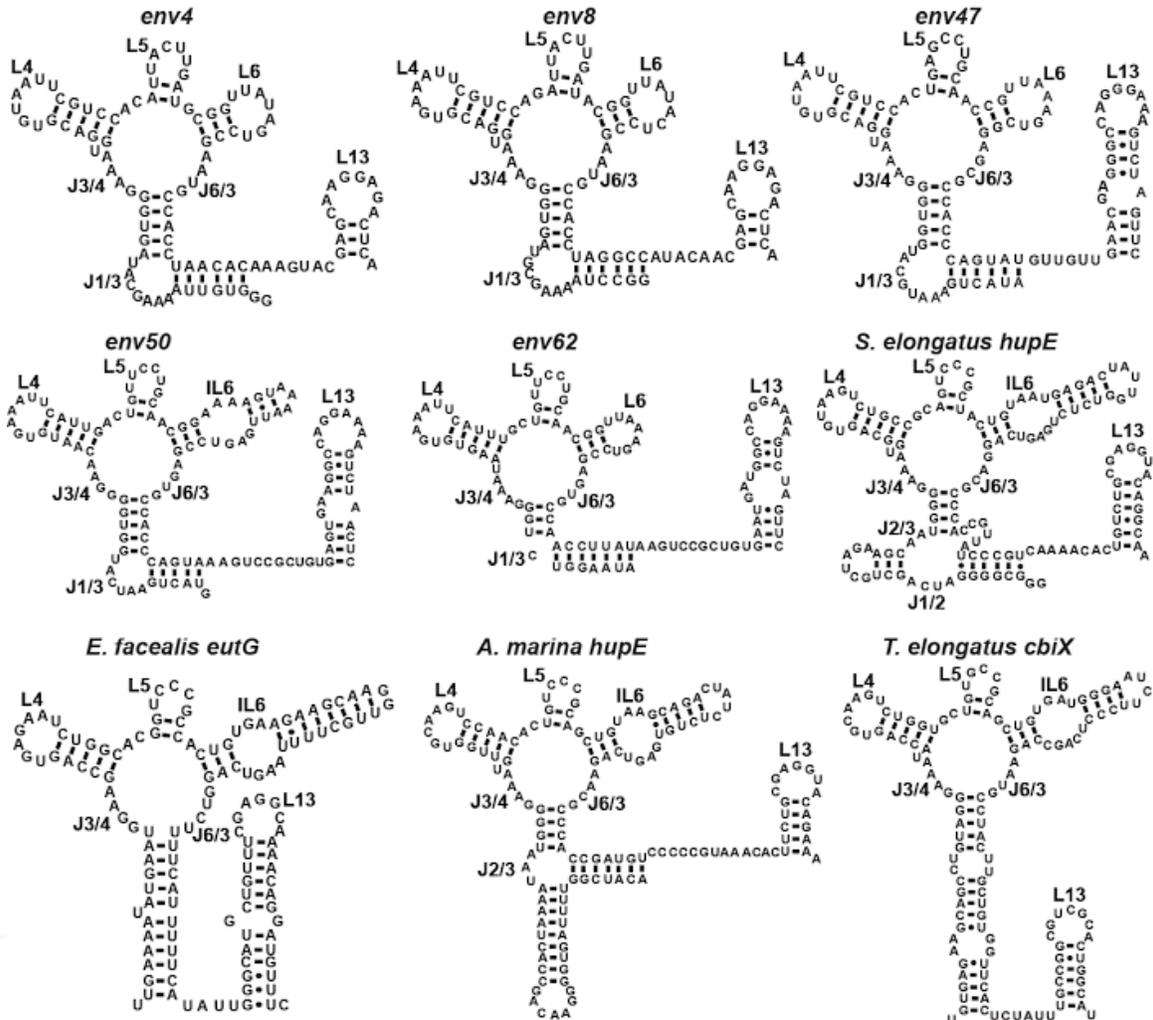
a.  $K_{rel} = (K_D, AdoCbl / K_D, MeCbl)$

b. These values taken from reference 32. All other values were measured by James Johnson Jr (see Acknowledgements)

c. These values are in singlet or are an average of duplicates. Other values are an average of triplicates ± the standard deviation.

This range in cobalamin affinity clearly shows that the initial hypothesis of universally MeCbl-specific Cbl-II riboswitches is incorrect. While the interaction of J1/3 and J6/3 assists in the steric occlusion of the 5'-dAdo moiety from the binding pocket in the *env8* riboswitch (**Figure 7**), this role of SD2 does not seem to be universal among Cbl-II riboswitches. Subdomain 2 varies greatly within the Cbl-II group; *env4*, *env8*, and

*env50* all contain large internal J1/3 bulges; *S. elongates* and *A. marina hupE* have three-way helical junctions; and *E. faecalis eutTG*, *T. elongates cbiX*, and *env62* have a near absence of any helical features in SD2 (Figure 8).



**Figure 8: Secondary structures of a sample of Cbi-II variants.** All sampled riboswitches have a central four-way junction, sequence patterns in L4 and L6 important to forming tertiary architecture, and sequences in L5 and L13 necessary for kissing loop formation. With the exception of *E. faecalis eutG*, J3/4 and J6/3 sequences have limited sequence diversity. SD2 architecture varies widely: *env4*, *env8*, and *env50* all have an internal bulge J1/3; *S. elongates* and *A. marina hupE* have three-way helical junctions; and *E. faecalis eutTG*, *T. elongates cbiX*, and *env62* have a near absence of any helical features. Figure taken from Jake Polaski.

The overall architecture of these riboswitches (**Figure 8**) is consistent with other cobalamin riboswitches. They include a central four-way junction which forms a key part of the ligand binding site, as well as conserved sequence patterns in L4 and L6 that are important to tertiary architecture organization, and sequences in L5 and L13 that are consistent with kissing loop formation (which allows ligand dependent gene expression) (**Figure 3A**). Additionally, with the exception of the *E. facealis eutTG* riboswitch, the bases in J3/4 and J6/3 that interact with the beta-axial moiety of cobalamin have limited sequence diversity; they conform to the consensus GRAA and RRYG, respectively. In the *E. facealis eutTG* riboswitch, J6/3 has uridine substitutions at the second and fourth position, which may confer its AdoCbl specificity in the absence of the IL J11/10 element.

Comparisons of Cbl-II riboswitch structures may offer some insight into the mechanism of Cbl-II cobalamin recognition, and the basis for non-selective binding. *Env62*, which has a central four-way junction that is nearly identical to *env4* and *env8*, but has no helical features in SD2, binds MeCbl and AdoCbl with the same affinity, while *env4* and *env8*, which contain an internal J1/3 bulge, are specific for MeCbl. *Env62*'s non-specific binding is similar to a variant of the *E. coli btuB* riboswitch that had the P8-P11 element deleted, and could no longer discriminate between species of cobalamin. This suggests that in the absence of peripheral elements, the conserved central binding core of Cbl-II riboswitches allows non-selective binding of all cobalamin species.

Further characterization of Cbl-II riboswitch selectivity is explored in this project. Site-directed mutagenesis is used to switch the selectivity of an AdoCbl-selective

riboswitch (WT*env50*) by substituting sequences from a MeCbl-specific riboswitch (WT*env8*). Mutations were designed with the binding pocket of the WT*env8* riboswitch in mind (**Figure 7**), and target the peripheral element J1/3 and single nucleotides in J3/4 and J6/3 of the central four-way junction, as well as the L4/L6 interaction, which organizes overall riboswitch architecture. By measuring relative binding affinities of the mutant riboswitch series to AdoCbl and MeCbl, we can extrapolate the roles of different elements in conferring ligand selectivity.

## 2. Methods

### 2.1 Construction of DNA Templates

DNA templates were constructed by recursive polymerase chain reaction (PCR)<sup>1</sup> from overlapping DNA oligonucleotides, using Pfu DNA polymerase. The PCR occurred in Pfu buffer (200 mM Tris, pH 8.8; 100 mM KCl, 1% Triton X-100, 100 mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>; 1 mg/mL BSA, 20 mM MgSO<sub>4</sub>), 2 mM dNTPs, 1 μM outside primers, and 10 nM inside primers, using 1:100 Pfu polymerase that was purified in house. The thermocycler protocol was as follows: 3 minutes at 90° C, then 30 cycles of 30 seconds at 90° C, 45 seconds at 55° C, and 45 seconds at 72° C. The annealing temperatures were adjusted for different constructs, in order to compensate for local hairpins or sequences of the oligonucleotides. Templates were constructed in 1 mL batches, for direct use in a subsequent transcription reaction. PCR templates were checked for the length of construct using 2% agarose gels and a low molecular weight DNA ladder.

### 2.2 Transcription of DNA Templates into RNA

---

<sup>1</sup> Unless otherwise stated, all basic RNA preparation and purification techniques used in the project (including PCR, in vitro transcription, and denaturing polyacrylamide gel electrophoresis) were performed as described in reference 39.

The 1 mL PCR reactions were used directly in the RNA transcription reaction. The 12.5 mL transcription reactions occurred in Transcription buffer (50 mM Tris-HCl, pH 7.5, 15 mM MgCl<sub>2</sub>, 5 mM dithiothreitol (DTT), 2 mM spermidine), with 32 mM MgCl<sub>2</sub>, 10 mM DTT, 4 mM rNTPs, the full 1 mL PCR template, 50 µL T7 RNA Polymerase, and 100 µL IPPase. The transcription reactions were incubated at 37° C for 3 hours, after which they were quenched with 30 mL 100% ethanol.

### *2.3 Purification of RNA*

The 12.5 mL RNA transcription reactions were quenched with ethanol, which also served to precipitate the RNA from the solution. The reactions were stored at -20° C overnight, and then spun down in a centrifuge to pellet the RNA. The ethanol was poured off and each pellet was re-suspended in 2 mL 8 M urea, 1 mL formamide, and 0.75 mL EDTA, and incubated at 65° C for 10 minutes to fully dissolve the pellet. 10 µL of 0.6% Bromophenol Blue was added to each tube. Two 12.5 mL RNA transcription reactions for each concept were ran on a single 8% denaturing acrylamide slab gel (a). The denaturing acrylamide was prepared from a 29:1 mix of acrylamide:bisacrylamide and had a final concentration of 8 M urea and 1x TBE buffer (89 mM Tris borate, 2mM EDTA, pH 8.2-8.4). The slab gels were run under electrophoresis at roughly 30 W for 4-6 hours, and then imaged using 254 nm (shortwave UV) light. The largest band of nucleic acid was excised from the gel and crushed and soaked in 0.5x TE buffer (10 mM Tris-HCl, 1mM EDTA, pH 8.0) overnight. The gel was then spun down and the supernatant was added to a 10K regenerated cellulose centrifugal filter and centrifuged; additional 0.5x TE buffer was added for three rounds of centrifugation to wash the RNA

and remove any residual urea. The washed RNA was run on a small denaturing gel with a small molecular weight RNA ladder in order to verify the size of the construct.

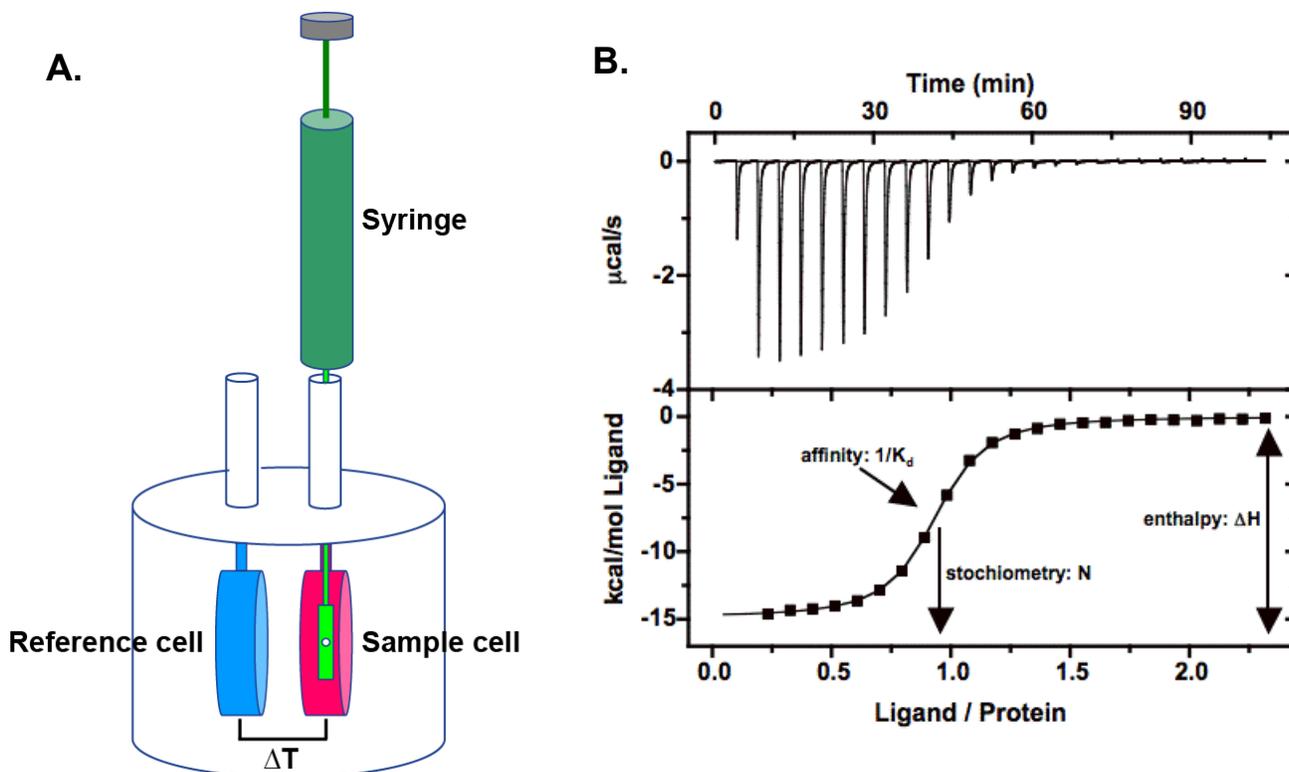
#### *2.4 Isothermal Titration Calorimetry*

The technique of isothermal titration calorimetry uses the direct measurement of heat generated or absorbed when molecules come into contact in order to determine thermodynamic and affinity parameters of the interaction (40). The technique was chosen for this project for a number of reasons. It is a true equilibrium method that only needs to be done at one temperature, and does not require any sort of molecular labeling. Fluorescent labeling of molecules in this project would have been problematic due to AdoCbl's tendency to photolyse (30, 31). The output of  $K_D$ ,  $\Delta H$ , and  $\Delta S$  from a single experiment was also appreciated.

The ITC calorimeter consists of a sample cell and a reference cell, surrounded by an adiabatic jacket, and a syringe that injects into the sample cell (**Figure 8A**) (40). The receptor molecule (in this case RNA) is loaded into the sample cell, while the ligand is loaded into the syringe. A small volume of ligand is injected into the sample cell, where the receptor and ligand can bind, in some cases giving off heat. The reference cell is kept at a constant reference power, and the Differential Power (DP, measured in  $\mu\text{cal/s}$ ) is the measured power differential needed to keep the sample cell and the reference cell at the same temperature. After each injection has equilibrated, the DP should return to the baseline, which is the reference power (**Figure 8B**). As more ligand is injected into the sample cell, the receptor molecule becomes saturated and less heat is released. (40)

The raw DP measurements are integrated (**Figure 8B**) to yield a plot of kcal/mol ligand vs. ligand/protein. A best-fit curve, which can be fit for a one-site or two-site binder, is fitted to the plot (40). This yields a number of parameters that characterize the molecular interaction. The inflection point in the curve is correlated to the N value, or the stoichiometry between the receptor and ligand. Based on crystallography data, Cbl-II riboswitches should all have an N value of 1. The slope of the curve at the inflection point is equal to  $1/K_D$  (the dissociation constant ( $K_D$ ) is a measure of affinity, and is equal to the concentration of ligand at 50% fraction bound.) Thermodynamic parameters can also be determined. Gibbs free energy,  $\Delta G$ , can be calculated using the  $K_D$  measurement; the enthalpy change,  $\Delta H$ , can be determined by the heat difference in the upper and lower baselines of the integrated plot; and the entropy,  $\Delta S$ , can be calculated using  $\Delta G$  and  $\Delta H$ . (40)

The quality and resolution of the data is dependent upon the relative values of the  $K_D$  and the concentrations of the ligand and receptor (40). This ratio is summed up in the c-value, which is equal to  $([R][L]/K_D)*N$ . With too low of a c-value ( $<1$ ), the transition is not well defined, and N,  $K_D$ , and  $\Delta H$  are not well resolved. With too high of a c-value ( $>1000$ ), the transition is too steep to resolve the  $K_D$  accurately, as the concentrations of ligand and receptor in the cell are too high relative to the  $K_D$ . A c-value between 5-500 is ideal. (40)

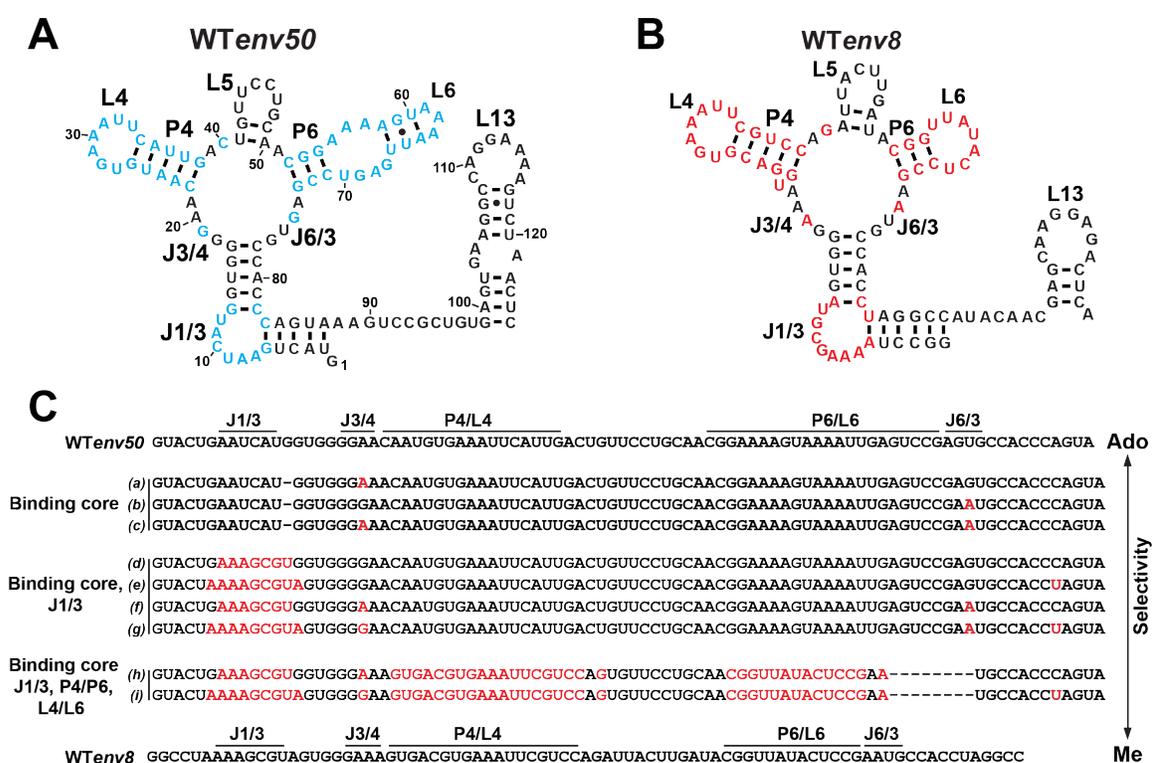


**Figure 8: ITC set-up schematic and sample trace.** A) ITC schematic. The receptor molecule is loaded into the sample cell, and the ligand molecule is loaded into the syringe, which carries out a series of injections. The reference cell is kept at a constant reference power, and the DP, or the power differential needed to keep the temperature difference between the reference and sample cell close to zero, is measured. Figure taken from reference 40. B) A sample ITC trace. The top graph shows raw DP data, in  $\mu\text{cal/sec}$  vs time in seconds, with each spike correlating to one injection. The bottom graph is a plot of the integrated spikes, with a best-fit curve defining the transition through the points. The inflection point correlates to the N value, or the stoichiometry between ligand and receptor molecules. The slope at the inflection point is equal to  $1/K_D$ , and the difference between the upper and lower baselines is the enthalpy change for the system upon binding. Figure taken from reference 41.

In this project, the RNA was dialyzed into an ITC buffer (42) (5 mM Na-MES, 100 mM KCl, 5 mM  $\text{MgCl}_2$ , pH 6.0) overnight using 6-8 kD MWCO dialysis membranes. The final concentration of the dialyzed RNA was taken using a Thermo Scientific NanoDrop 2000c Spectrophotometer. The stock RNA was aliquoted into appropriate dilutions (depending on the expected  $K_d$  with MeCbl and AdoCbl). Dilutions of either MeCbl or AdoCbl were made to match 10x the expected RNA concentration. Before loading the



These same regions became a target for mutagenesis in the WT*env50* riboswitch. Hybrid mutants were designed by systematically substituting WT*env8* sequences (Figure 9B) into analogous regions in WT*env50* (Figure 9A), in the hopes of shifting the AdoCbl-selective riboswitch toward MeCbl-selectivity. The P4/P6 interaction (Figure 10, top, black) was also targeted because it differs so drastically between WT*env8* and WT*env50*, and is important for overall riboswitch architecture organization. A summary of the mutant sequences is shown in Figure 9C.



**Figure 9: WT*env50* and WT*env8* riboswitches and a sequence summary of directed mutations.** A) Secondary structure of WT*env50* riboswitch. Mutation targets are in blue. B) Secondary structure of WT*env8* riboswitch. Mutation targets are in red. C) Sequence summary of directed mutations to WT*env50* riboswitch. **Binding core** mutations (*env50a,b,c*) are directed at bases in J3/4 and J6/3 that differ between WT*env50* and WT*env8*. Combined **Binding core, J1/3** mutations (*env50d,e,f,g*) are directed at the same bases in J3/4 and J6/3, and WT*env50* J1/3 is replaced with the WT*env8* J1/3, with varying bases at the neck of the junction. **Binding core, J1/3, P4/P6, and L4/L6** mutations (*env50h,i*) have additional mutations in the P4/L4 and P6/L6 peripheral elements, replacing the WT*env50* extensions with the WT*env8* extensions. Mutated bases are shown in red. Figure taken from Jake Polaski.

### 3.2 Affinities of env50 Mutants for AdoCbl and MeCbl

The *env50* mutants were tested via ITC with both AdoCbl and MeCbl compounds. The dissociation constant ( $K_D$ ) for both ligands was measured, and a  $K_{rel}$  ( $K_{D\ Ado}/K_{D\ Me}$ ) was calculated in order to show relative selectivity (**Table 2**). WT*env50* and WT*env8* binding affinities are included for the sake of comparison.

| <i>riboswitch</i>            | $K_D$ , AdoCbl (nM) | $K_D$ , MeCbl (nM) | $K_{rel}^c$ |
|------------------------------|---------------------|--------------------|-------------|
| WT <i>env50</i> <sup>d</sup> | 2 ± 0.58            | 190 ± 23           | 0.011       |
| <i>env50a</i> <sup>e</sup>   | 29 <sup>a</sup>     | 97 ± 42            | 0.30        |
| <i>env50b</i> <sup>e</sup>   | 30 <sup>a</sup>     | 61 <sup>a</sup>    | 0.49        |
| <i>env50c</i> <sup>e</sup>   | 36 ± 4.5            | 30 <sup>a</sup>    | 1.2         |
| <i>env50d</i>                | 4.6 ± 3.1           | 15 ± 13            | 0.31        |
| <i>env50e</i>                | 8.5 ± 4.9           | 6.2 <sup>a</sup>   | 1.4         |
| <i>env50f</i>                | 72 <sup>a</sup>     | 2.2 ± 0.49         | 33          |
| <i>env50g</i>                | 140 ± 30            | <3.9 <sup>b</sup>  | >36         |
| <i>env50h</i>                | 81 ± 51             | 6.1 ± 0.6          | 13          |
| <i>env50i</i>                | 270 ± 64            | <4.0 <sup>b</sup>  | >68         |
| WT <i>env8</i> <sup>d</sup>  | >50000              | 7.5 ± 4            | >6600       |

**Table 2: Affinities of Cbl-II variants and mutants for AdoCbl and MeCbl.**

a. These values are an average of duplicates. Other values are averages of triplicates ± standard deviation.

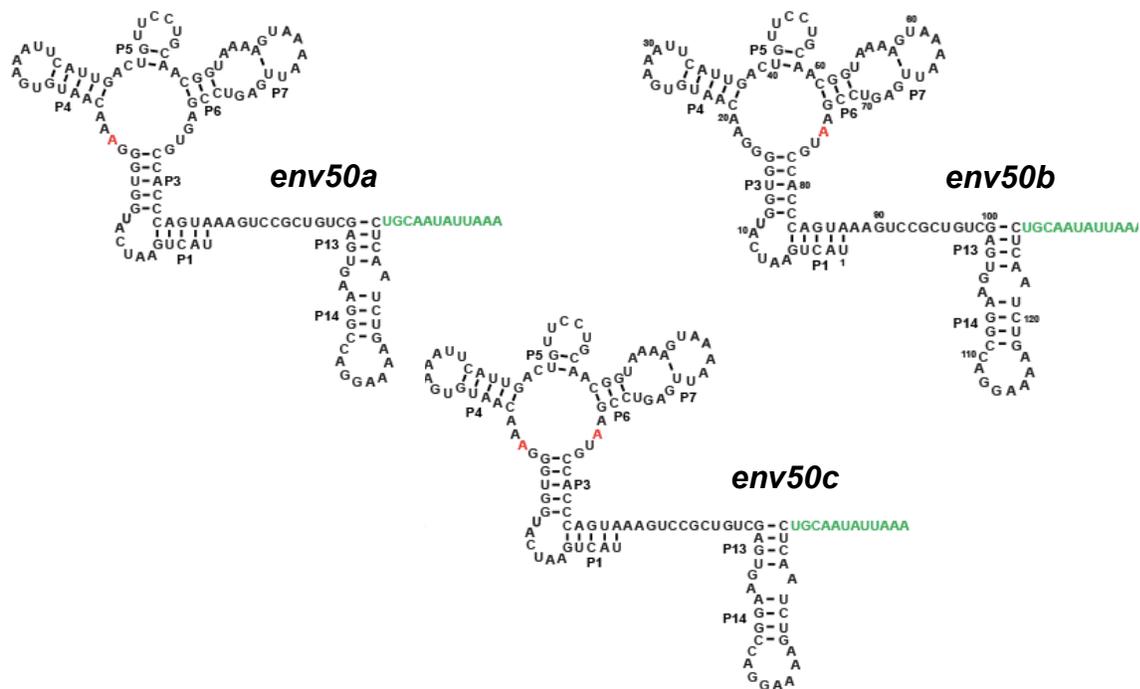
b. These values are averages of triplicates with high c-values (>1000). They are presented as less than the average because both had  $K_D$ 's measured in pM and were probably not resolved to the true  $K_D$ .

c.  $K_{rel} = (K_{D\ Ado}/K_{D\ Me})$  d. Values from reference 32. e. Values measured by James Johnson Jr (see Acknowledgements)

### 3.3 Effect of env50 Mutations on Ligand Selectivity

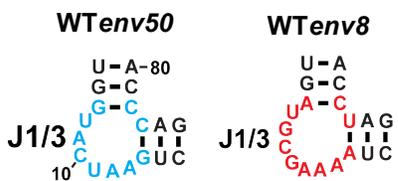
The first mutations were made to binding core bases that have direct contact with the cobalamin beta-axial moiety, and served to make the RNA weakly selective. The

sequences of J3/4 and J6/3 vary at only two positions between WT*env50* and WT*env8*: G19 and G75 in WT*env50* (**Figure 9A**) are adenosines in WT*env8* (**Figure 9B**). Mutation of G19→A or G75→A (*env50a* and *env50b* mutants, **Figure 10**) decreased AdoCbl affinity by >10-fold (**Table 2**), but only moderately increased MeCbl affinity, resulting in riboswitches with a slight (2 to 3-fold) selectivity for AdoCbl. When both G19A and G75A were mutated (*env50c*, **Figure 10**), the resulting riboswitch could not discriminate between the species of cobalamin, though it had a reasonably high affinity for both AdoCbl and MeCbl. Importantly, while these mutations resulted in a loss of function with the decreased binding of AdoCbl, they also resulted in a gain of function with the increased binding of MeCbl. All mutants have activity, which indicates that the mutations are relevant to riboswitch function and do not merely disrupt functional structure.



**Figure 10: Mutations in the J3/4 and J6/3 binding core.** *Env50a* has a G19A mutation in J3/4; *env50b* has a G75A mutation in J6/3; *env50c* has both. These bases have been shown to interact directly with the beta-axial moiety of cobalamin. Mutated bases are shown in red. Green bases were included in mutant sequences to fully encapsulate the 3' end of the riboswitch. Figure taken from Jake Polaski.

The failure of the J6/3 and J3/4 mutations to switch selectivity shows that the binding core alone cannot fully establish Cbl-II selectivity. The next target for mutation was the J1/3 peripheral element, and mutants were made with the entire J1/3 junction of WT*env8* substituted into the *env50* scaffold. The actual sequences of WT*env8* J1/3 (Figure 11, red) and WT*env50* J1/3 (Figure 11, blue) contain some similarities; they only differ by one base in length, and include a U (U8 in WT*env50*), a C (C10), and two A's (A12-A13) at the same positions. However, the pairs that make up the neck of the junction are an A-U base pair and an A-C mismatch in WT*env8*, and two G-C base pairs in WT*env50*. The replacement of A9 with a G, and G11 with a U, as well as the addition of an A after A12 also contribute to structural differences between the J1/3 elements.



**Figure 11: J1/3 element in WT*env50* and WT*env8*.**

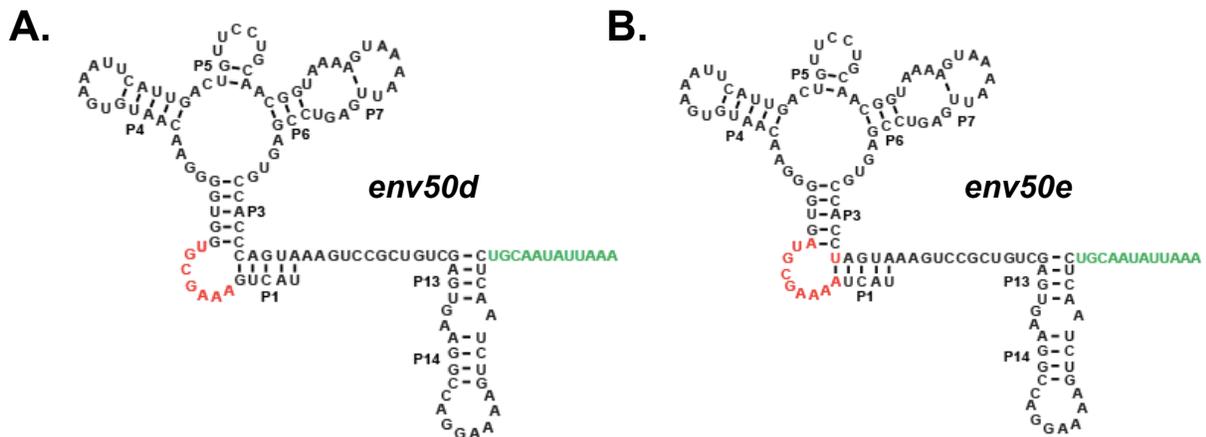
WT*env50* is shown in blue; WT*env8* is shown in red. Bases differ at the neck region, and WT*env50* A9 and U11. WT*env8* also has an additional A after A13. Figure taken from Jake Polaski.

There were two variations on the J1/3

substitution; one mutant keeps the two G-C base pairs at the neck of the junction that are present in WT*env50* (*env50d*, Figure 12A) while the other substitutes the A-U pair and A-C mismatch from WT*env8* (*env50e*, Figure 12B). In both mutants, there was a diminished selectivity for AdoCbl (2 to 3-fold), as well as a significantly increased affinity

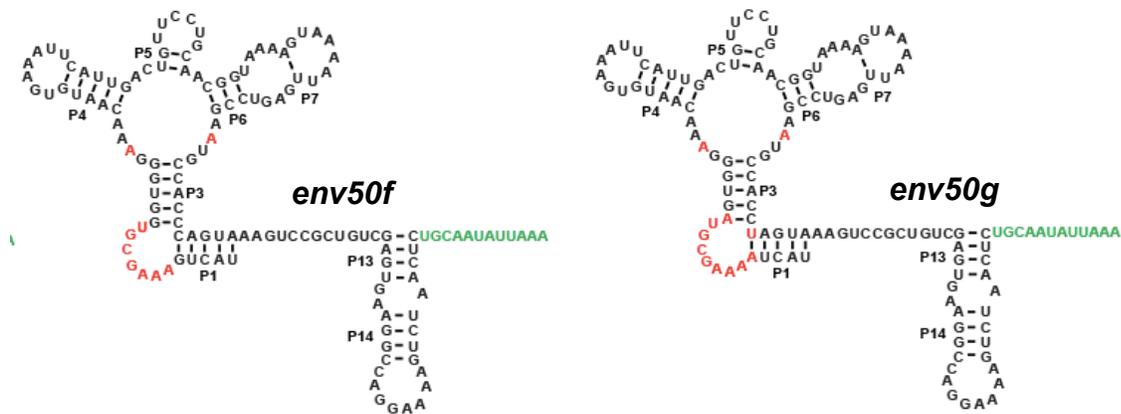
for MeCbl (10-fold). The *env50d* riboswitch overall still had a 3-fold preference for AdoCbl, but the neck mutations in *env50e* shifted the riboswitch into a slight MeCbl preference; this trend continues in later mutants with J1/3 substitutions (*env50f-i*). The relative flexibility of the AC-AU neck compared to the GC-GC clamp and the resulting expansion of conformational sampling by the riboswitch may have something to do with

this disparity (43). J1/3 clearly has some effect on the binding pocket, but its impact on ligand preference was about the same as the core mutations. On its own, J1/3 cannot fully establish Cbl-II selectivity.



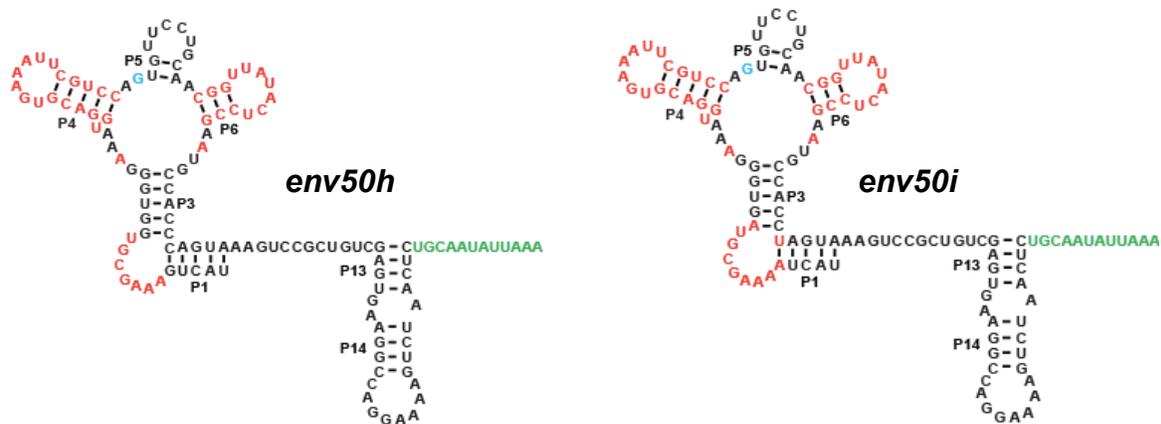
**Figure 12: Mutations to peripheral element J1/3.** The WT *env8* J1/3 element was substituted into the *env50* structure. Two variations of the substitution were made: A) the two WT *env50* G-C base pairs at the neck of J1/3 were conserved, and B) the WT *env8* A-U pair and A-C mismatch were substituted. Mutated bases are shown in red. Green bases were included in mutant sequences to fully encapsulate the 3' end of the riboswitch. Figure taken from Jake Polaski.

The combination of binding core mutations and J1/3 substitution resulted in riboswitches (*env50f* and *env50g*, **Figure 13**) with substantial MeCbl selectivity. In the case of *env50g*, the RNA had a 36-fold MeCbl preference. That the ligand preference switch came from a combination of core and J1/3 mutations, when each mutation on its own only moved the RNA into a non-specific binding range, is evidence that the interaction of J1/3 and J6/3 is crucial to conferring ligand selectivity.



**Figure 13: Combination of binding core and J1/3 mutations.** Both A's are mutated in J3/4 and J6/3; both variations of the J1/3 substitution are made: *env50f* follows **Figure 11A** and *env50g* follows **Figure 11B**. Mutated bases are shown in red. Green bases were included in mutant sequences to fully encapsulate the 3' end of the riboswitch. Figure taken from Jake Polaski.

Mutations to peripheral elements P4/L4 and P6/L6, in addition to the binding core and J1/3 mutations (*env50h* and *env50i*, **Figure 14**), resulted in only a moderate change in ligand preference by the riboswitches. (Both mutants retained MeCbl preference, though *env50h* became half as selective for MeCbl while *env50i* became twice as selective.) Though the L4/L6 interaction is important for tertiary structure organization, it does not appear to substantially affect ligand selectivity. The presence of the L4 and L6 peripheral elements are likely essential for general binding, since they help to organize the tertiary architecture of the RNA, but are not critical to selectivity.



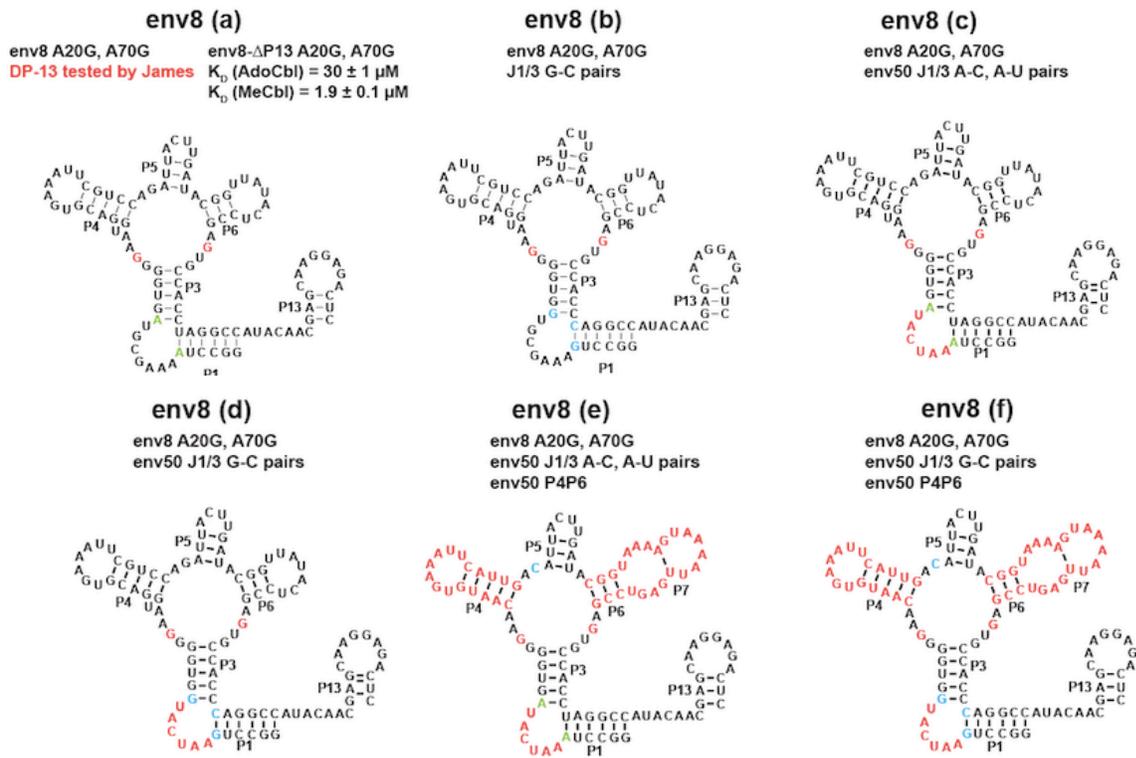
**Figure 14: Mutations to binding core, J1/3, and P4/L4 and P6/L6.** The same mutations as in **Figure 12**, with the addition of WT*env8* P4/L4 and P6/L6 substitutions into *env50*. The L4/L6 interaction is important for riboswitch tertiary structure organization. Mutated bases are shown in red. Green bases were included in mutant sequences to fully encapsulate the 3' end of the riboswitch. Figure taken from Jake Polaski.

While the project was successful in shifting ligand preference from AdoCbl to MeCbl, the specificity of the most selective mutant riboswitch falls far short of WT*env8*, which has a greater than 6600-fold preference for MeCbl. This final step in conferring ligand specificity, rather than mere ligand selectivity, remains elusive. Since *env50i* (the most MeCbl-selective of the mutants) is structurally extremely similar to WT*env8* (with the same binding core, P4, P6, and J1/3), this gap in specificity is somewhat surprising. It's possible that the interaction of L5 and L13 (which makes up the translational regulatory domain, and does not come into contact with the binding pocket) also has something to do with ligand selectivity. L13 in WT*env50* looks quite different from L13 in WT*env8* (**Figure 9A,B**); it is much longer and has an additional internal loop. The contributions of this element to cobalamin species recognition have not been explored.

### 3.4 Current Experiments with *env8* Mutants

Given the success in switching the WT*env50* riboswitch from AdoCbl-selective to MeCbl-selective, it seems the logical next step would be to attempt to switch the

WT*env8* riboswitch from MeCbl-specific to AdoCbl-selective. Employing the same strategy of mutating sequences to replicate pieces of WT*env50* architecture, we have designed a series of *env8* mutants (**Figure 15**) that target the same elements as in the *env50* series, including the disparate binding core bases, the J1/3 peripheral element, and the P4/P6 interaction.



**Figure 15: A series of *env8* mutants with WT*env50* elements.** Mutation targets include binding core bases (*env8a* and *env8b*), binding core + J1/3 (*env8c* and *env8d*), and binding core + J1/3 + P4/P6 (*env8e* and *env8f*). Mutated bases are shown in red. Figure taken from Jake Polaski.

This set of experiments will hopefully give us additional insight into WT*env8* MeCbl binding, and also into the unusual case of WT*env50*'s general (though AdoCbl-selective) binding of both MeCbl and AdoCbl. It would be ideal to have a crystal structure of WT*env50* bound to AdoCbl in order to fully discern the placement of the

binding core and J1/3 in relation to the beta-axial moiety. However, X-ray crystallography is notoriously difficult with AdoCbl, since every step must take place in the dark in order to avoid photolysis, and the very act of shooting the crystals can destroy the compound (32).

ITC experiments on this set of mutants have begun, and preliminary data is included in **Table 3**.

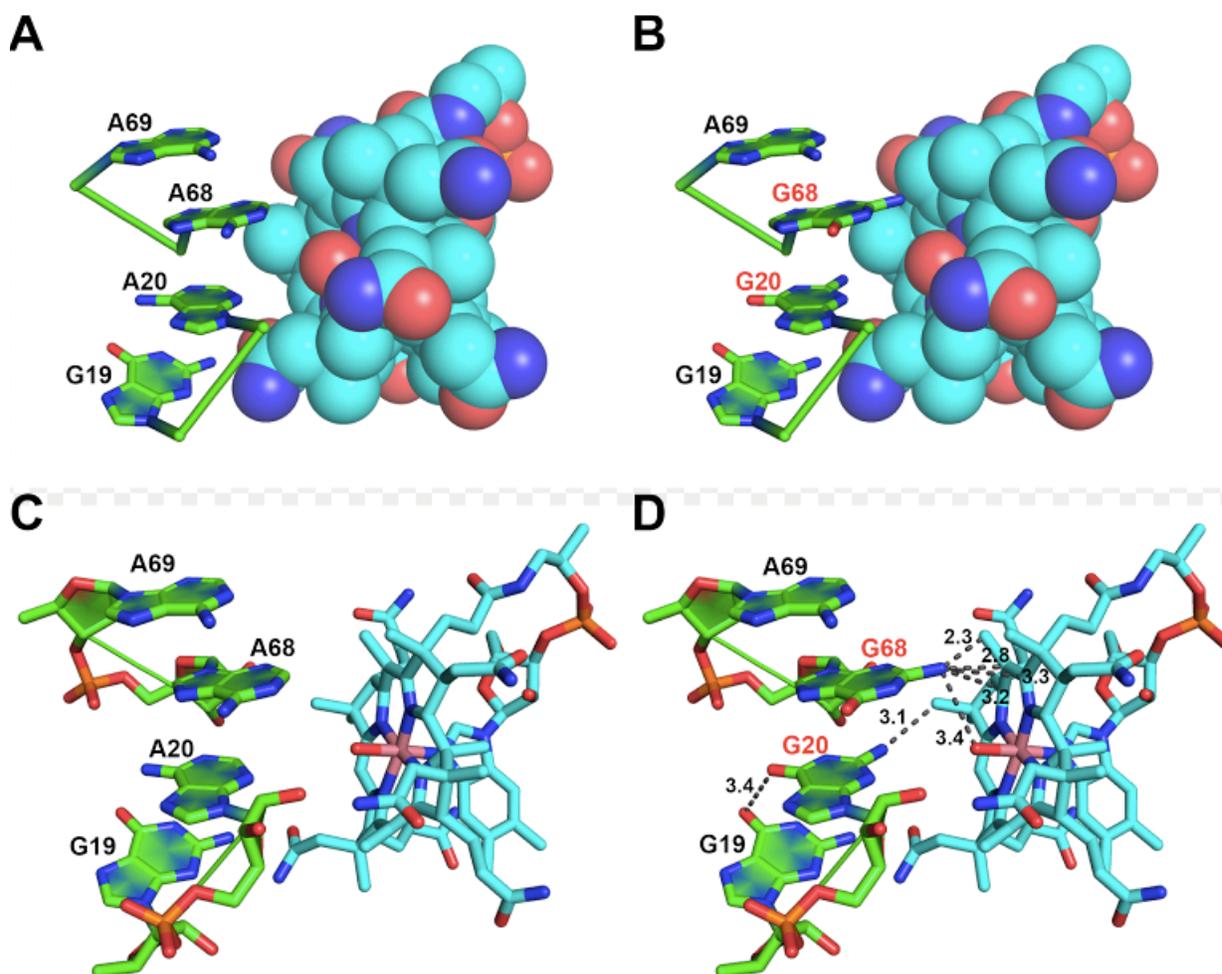
| <i>riboswitch</i>     | $K_D$ , AdoCbl (nM) | $K_D$ , MeCbl (nM) | $K_{rel}^b$ |
|-----------------------|---------------------|--------------------|-------------|
| <b>WT<i>env8</i></b>  | >50000              | $7.5 \pm 4$        | >6600       |
| <b><i>env8a</i></b>   | 2700 <sup>a</sup>   | 2400 <sup>a</sup>  | 0.89        |
| <b><i>env8b</i></b>   | 79000 <sup>a</sup>  | --                 | --          |
| <b><i>env8c</i></b>   | --                  | --                 | --          |
| <b><i>env8d</i></b>   | 62 <sup>a</sup>     | $5900 \pm 1100$    | 0.010       |
| <b><i>env8e</i></b>   | --                  | $110 \pm 25$       | --          |
| <b><i>env8f</i></b>   | $19 \pm 5$          | $2500 \pm 530$     | 0.0075      |
| <b>WT<i>env50</i></b> | $2 \pm 0.58$        | $190 \pm 23$       | 0.011       |

**Table 3: Preliminary data for *env8* mutant series.**

a. These values are an average of duplicates or a singlet value. Other values are averages of triplicates  $\pm$  standard deviation. Dashes indicate data that has not yet been collected or could not yet be calculated.

b.  $K_{rel} = (K_{D\ Ado}/K_{D\ Me})$

It does appear that in *env8d* and *env8f*, we have managed to switch the selectivity of the RNA (and indeed created riboswitches that are more AdoCbl-selective than WT*env50*.) Further experiments will have to be conducted to determine the components of the riboswitch most responsible for the switch, but there is ongoing speculation, particularly concerning the differences in the WT*env8* and WT*env50* binding cores.



**Figure 16: Binding core interactions with beta-axial moiety of cobalamin.** A) Space filling structure of cobalamin interacting with A69, A68, A20, and G19 of *WTenv8*. The beta-axial moiety is in the middle of the cobalamin in red. B) Same as in A, but with A68 and A20 replaced with G's to replicate the binding core of *WTenv50*. C) Same as A, but with a ball and stick model of cobalamin. D) Same as B, but with a ball and stick model of cobalamin. Contact distances between chemical groups are shown in angstroms. Figure taken from Jake Polaski.

In **Figure 16**, the *WTenv8* crystal structure was used to model a speculative *WTenv50* binding core. Bases A68 and A20 were replaced by G's, and the resulting contact distances between chemical groups in the RNA and the cobalamin molecule were measured (**Figure 16D**). The model shows that the replacement of A with a G at position 68 introduces a great deal of steric hindrance between the amino group of the guanine and several pieces of the cobalamin, including the beta-axial moiety (shown

extending from the middle of the cobalamin, in red). Given this, and the steric interactions of G20 with G19 and the cobalamin molecule, it is unlikely these bases would be positioned in the same way in the WT*env50* binding core. We posit that the inclusion of these nucleotides in the J3/4 and J6/3 elements of WT*env50* causes a natural expansion of the binding core, so that these bases do not pack quite as tightly against the cobalamin face as they do in WT*env8*. This would allow WT*env50* to accommodate the bulkier AdoCbl more easily into the binding pocket. ITC experiments measuring the effect of binding core base mutations on affinity are in progress.

### 3.5 The Bigger Picture of Non-Specific Ligand Binding

A range of RNA aptamers use different strategies to recognize their ligands, which result in varying degrees of specificity. *T. maritima lysC* riboswitch has a rigid binding pocket (21), which requires lysine to fulfill H-bonds from main chain atoms in the amino acid backbone, as well as an electrostatic interaction between lysine's epsilon-amino group and the RNA backbone. Since lysine is the only amino acid that can fulfill both, the riboswitch is highly specific (>10,000 fold over other amino acids) (19, 21, 22). Other riboswitches, like the tetrahydrofolate (THF) riboswitch, need to recognize a group of related ligands in a cell (44, 45). In order to effectively monitor the cellular folate pool, the riboswitch needs to recognize all folates with a reduced pterin moiety, but discriminate against fully oxidized folates. In bacteria the pool of reduced folates is chemically heterogeneous due to diverse one-carbon groups on the N5/N10 atoms and variable polyglutanylation, which complicates the task of recognizing the whole set. The RNA solves this problem by ignoring the variable regions, and only recognizes the face of the pterin moiety that does not carry one-carbon groups (44, 45).

A subset of Cbl-II WT variants are able to recognize chemically distinct cobalamin forms. Many of these variants came from a surface Pacific Ocean metagenome; it's possible that their source organisms need to recognize all species of cobalamin, as the AdoCbl pool builds up at night then photolyses during the day. Unlike the THF riboswitch, these RNAs don't have the luxury of ignoring variable groups because the beta-axial moiety of cobalamin juts directly into the binding core. What, then, is the structural basis for this non-specific binding of chemically distinct metabolites? From this study, it is clear that the interaction of J1/3 and J6/3 are critical to the occlusion or permission of the 5'-dAdo moiety to the binding pocket. In the absence of J1/3, as in *env62*, binding is non-selective, indicating that the central four-way junction alone allows for general binding. The interaction of WT *env50*'s J1/3 with the J6/3 binding region when the riboswitch is bound to either MeCbl or AdoCbl is yet unknown, as is the packing of binding core bases against the beta-axial moiety, and structural knowledge of the riboswitch could provide insight into a novel mechanism of general ligand binding. Other riboswitch aptamers, such as the *xpt/pbuX* guanine binding riboswitch, have been shown to use local conformational flexibility within their binding pockets, rather than repositioning of the ligand, to accommodate a variety of purine analogues (46). It's possible that non-specific Cbl-II variants employ a similar strategy.

## References

1. Hangauer MJ, Vaughn IW, McManus MT. (2013) Pervasive Transcription of the Human Genome Produces Thousands of Previously Unidentified Long Intergenic Noncoding RNAs. *PLOS Genetics* 9(6), e1003569
2. Cech TR, Steitz JA. (2014) The noncoding RNA revolution-trashing old rules to forge new ones. *Cell* 157(1), 77-94.
3. Lee RC, Feinbaum, RL, and Ambros, V. (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75, 843–854
4. Wightman B, Ha I, Ruvkun G. (1993) Posttranscriptional regulation of the heterochronic gene *lin-14* by *lin-4* mediates temporal pattern formation in *C. elegans*. *Cell*. 75, 855–862
5. Oxender DL, Zurawski G, Yanofsky C. (1979) Attenuation in the *Escherichia coli* tryptophan operon: role of RNA secondary structure involving the tryptophan codon region. *PNAS* 11, 5524-8.
6. Nahvi A, Sudarsan N, Ebert MS, Zou X, Brown KL, Breaker RR (2002). Genetic control by a metabolite binding mRNA. *Chem Biol* 9 (9): 1043–1049.
7. Breaker RR. (2012) Riboswitches and the RNA world. *Cold Spring Harbor Perspect Biol* 4(2), a003566.
8. Garst AD, Edwards AL, Batey RT. (2011) Riboswitches: structures and mechanisms. *Cold Spring Harbor Perspect Biol* 3(6), a003533.
9. Garst AD, 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.
10. Irnov, Kertsburg A, Winkler WC. (2006) Genetic control by cis-acting regulatory RNAs in *Bacillus subtilis*: general principles and prospects for discovery. *Cold Spring Harb Symp Quant Biol*. 71, 239-49.
11. Mulhbacher J, Lafontaine DA. (2007) Ligand recognition determinants of guanine riboswitches. *Nucleic Acids Res*. 35(16), 5568-80.
12. Barrick JE, Breaker RR. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol*. 8, R239.
13. Mandal M, Breaker RR. (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol*. 11(1), 29-35.

14. Winkler W, Nahvi A, Breaker RR. (2002) Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression. *Nature*. 419(6910), 952-6.
15. Kim JN, Breaker RR. (2008) Purine sensing by riboswitches. *Biol Cell* 100(1), 1-11.
16. Batey RT. (2011) Recognition of S-adenosylmethionine by riboswitches. *Wiley interdisciplinary reviews. RNA* 2, 299-311.
17. Batey RT. (2012) Structure and mechanism of purine-binding riboswitches. *Quarterly reviews of biophysics* 45, 345-381.
18. Serganov A. (2009) The long and the short of riboswitches. *Curr Opin Struct Biol* 19, 251-259.
19. Garst AD, Porter EB, Batey RT. (2012) Insights into the regulatory landscape of the lysine riboswitch. *J Mol Biol* 423, 17-33.
20. Sudarsan N, Wickiser JK, Nakamura S, Ebert MS, Breaker RR. (2003) An mRNA structure in bacteria that controls gene expression by binding lysine. *Genes & development* 17, 2688-2697.
21. Garst AD, Heroux A, Rambo RP, Batey RT. (2008) Crystal structure of the lysine riboswitch regulatory mRNA element. *The Journal of biological chemistry* 283, 22347-22351.
22. Serganov A, Huang L, Patel DJ. (2008) Structural insights into amino acid binding and gene control by a lysine riboswitch. *Nature* 455, 1263-1267.
23. Nahvi A, Barrick JE, Breaker RR. (2004) Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes. *Nucleic Acids Res.* 32, 143-150.
24. Lundrigan MD, Koster W, Kadner RJ. (1991) Transcribed sequences of the *Escherichia coli btuB* gene control its expression and regulation by vitamin B12. *Proc Natl Acad Sci USA* 88, 1479-1483.
25. Franklund CV, Kadner RJ. (1997) Multiple transcribed elements control expression of the *Escherichia coli btuB* gene. *J Bacteriol.* 179, 4039-4042.
26. Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS. (2003) Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element. *RNA*. 9, 1084-1097.

27. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. (2003) Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *J Biol Chem.* 278, 41148–41159.
28. Bucher D, Sandala GM, Durbeej B, Radom L, Smith DM. (2012) The elusive 5'-deoxyadenosyl radical in coenzyme-B12-mediated reactions. *J Am Chem Soc.* 134(3), 1591-9.
29. Banerjee R and Ragsdale SW. (2003) The many faces of vitamin B12: catalysis by cobalamin-dependent enzymes. *Annual review of biochemistry* 72, 209-247.
30. Baxter N, Horsford J, Wokes F, Norris F, Fernandes S. (1953) Cyanocobalamin and hydroxocobalamin in vitamer B12 injections. *Journal of Pharmacy and Pharmacology* 5, 723-736.
31. Veer W, Edelhausen J, Wijmenga H, Lens J. (1951) Vitamin B12: I. The relation between vitamins B12 and B12b. *Biochimica et biophysica acta* 6, 225-228.
32. Johnson JE Jr, Reyes FE, Polaski JT, Batey RT. (2012) B12 cofactors directly stabilize an mRNA regulatory switch. *Nature* 492, 133-137.
33. Peselis A, Serganov A. (2012) Structural insights into ligand binding and gene expression control by an adenosylcobalamin riboswitch. *Nature structural & molecular biology* 19, 1182-1184.
34. Choudhary PK, Sigel RK. (2013) Mg<sup>2+</sup>-induced conformational changes in the btuB riboswitch from *E. coli*. *RNA* 20(1), 36-45.
35. Gallo S, Mundwiler S, Alberto R, Sigel RK. (2011) The change of corrin-amides to carboxylates leads to altered structures of the B12-responding btuB riboswitch. *Chemical communications* 47, 403-405.
36. Weinberg Z, Wang JX, Bogue J, Yang J, Corbino K, Moy RH, Breaker RR. (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome biology* 11, R31.
37. Fox KA, Ramesh A, Stearns JE, Bourgoigne A, Reyes-Jara A, Winkler WC, Garsin DA. (2009) Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in *Enterococcus faecalis*. *Proceedings of the National Academy of Sciences of the United States of America* 106, 4435-4440.
38. DebRoy S, Gebbie M, Ramesh A, Goodson JR, Cruz MR, van Hoof A, Winkler WC, Garsin DA. (2014) A riboswitch-containing sRNA controls gene expression by sequestration of a response regulator. *Science* 345, 937-940.

39. Edwards AL, Garst AD, Batey RT. (2009) Determining structures of RNA aptamers and riboswitches by X-ray crystallography. *Methods Mol Biol.* 535, 135-63
40. Milev, S. (2013). *Isothermal titration calorimetry: Principles and experimental design.* GE Healthcare.
41. Tu H. Isothermal Titration Calorimeter. (2015). Biocenter Oulu. Retrieved from <http://www oulu.fi/biocenter/biophysical-protein-analysis/itc>
42. Gilbert SD, Batey RT. (2009) Monitoring RNA-ligand interactions using isothermal titration calorimetry. *Methods Mol Biol.* 540, 97-114.
43. Vicens Q, Mondragón E, Batey RT. (2011) Molecular sensing by the aptamer domain of the FMN riboswitch: a general model for ligand binding by conformational selection. *Nucleic Acids Res.* 39(19), 8586-98.
44. 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.
45. Ames TD, Rodionov DA, Weinberg Z, Breaker RR. (2010) A Eubacterial Riboswitch Class that Senses the Coenzyme Tetrahydrofolate. *Chem Biol.* 17(7), 681-5.
46. Gilbert SD, Reyes FE, Edwards AL, Batey RT. (2009). Adaptive ligand binding by the purine riboswitch in the recognition of guanine and adenine analogs. *Structure* 17(6), 857-68.
47. Croft MT, Lawrence AD, Raux-Deery E, Warren MJ, Smith AG. (2005). Algae acquire vitamin B<sub>12</sub> through a symbiotic relationship with bacteria. *Nature* 438, 90-93.
48. Bradbeer C, Woodrow ML. (1976). Transport of Vitamin B<sub>12</sub> in *Escherichia coli*: Energy Dependence. *Journal of Bacteriology* 128(1), 99-104.

## **Acknowledgements**

I'd like to thank the entire Batey lab, particularly Dr. Rob Batey, who gave me such a wonderful opportunity to explore my love of science in a real and tangible way, and who sponsored me financially during the summers, so I could continue to do work I really loved. A big shout out goes to Jake Polaski, who has been a standout mentor and "lab dad" for the whole project. His patience and guidance through the process has been invaluable, and his organizational skills are to be envied. Hopefully when I'm a grad student, I'll be just like him. Bagde Akdobar, Zac Holmes, Michal Matyjasik, Antonia Lin, and Lea Drogalis have all conspired to make lab both a fun and productive place, and I'm glad to count them as friends. I'd also like to thank past mentors, Ely Porter and Joan Marcano-Velásquez, who had the trouble of teaching a more naïve Sam how to pipette and also how to properly organize and label dozens of simultaneous PCR reactions. I have never met James Johnson Jr, but his lab notebook is the best I've ever seen and I'm grateful that it was so easy to mine his data. The work discussed in this thesis was performed as a collaboration between myself, James Johnson Jr, and Jacob Polaski. Various aspects of binding titrations were performed by all parties.