9094–9105 *Nucleic Acids Research, 2018, Vol. 46, No. 17* *Published online 26 June 2018*
doi: 10.1093/nar/gky539

A functional genetic screen reveals sequence preferences within a key tertiary interaction in cobalamin riboswitches required for ligand selectivity

Jacob T. Polaski, Otto A. Kletzien, Lea K. Drogalis and Robert T. Batey*

Department of Chemistry and Biochemistry, University of Colorado, Boulder, CO 80309, USA

Received April 10, 2018; Revised May 25, 2018; Editorial Decision May 29, 2018; Accepted May 30, 2018

ABSTRACT

Riboswitches are structured mRNA sequences that regulate gene expression by directly binding intracellular metabolites. Generating the appropriate regulatory response requires the RNA rapidly and stably acquire higher-order structure to form the binding pocket, bind the appropriate effector molecule and undergo a structural transition to inform the expression machinery. These requirements place riboswitches under strong kinetic constraints, likely restricting the sequence space accessible by recurrent structural modules such as the kink turn and the T-loop. Class-II cobalamin riboswitches contain two T-loop modules: one directing global folding of the RNA and another buttressing the ligand binding pocket. While the T-loop module directing folding is highly conserved, the T-loop associated with binding is substantially less so, with no clear consensus sequence. To further understand the functional role of the binding-associated module, a functional genetic screen of a library of riboswitches with the T-loop and its interacting nucleotides was used to build an experimental phylogeny comprised of sequences that possess a wide range of cobalamin-dependent regulatory activity. Our results reveal conservation patterns of the T-loop and its interaction with the binding core that allow for rapid tertiary structure formation and demonstrate its importance for generating strong ligand-dependent repression of mRNA expression.

INTRODUCTION

Riboswitches are regulatory RNA elements found in the 5′-leader of bacterial mRNAs that modulate gene expression via their interaction with a small molecule metabolite. Regulatory activity is achieved through communication between two functional domains: a receptor (aptamer) domain that interrogates the local environment for the cognate metabolite and a regulatory (expression platform) domain that undergoes a structural transition based on the occupancy status of the receptor to direct the gene expression machinery (1,2). Currently, much of the information about the relationship between folding and regulatory activity of riboswitches has been inferred from *in vitro* studies of isolated receptor domains, resulting in a limited understanding of how conformational alterations drive regulatory activity in a cellular environment. Developing a more complete understanding of this relationship is important, as gene regulation under biological conditions places riboswitches under strong kinetic constraints requiring the aptamer domain to rapidly fold into a functional structure with high fidelity, limiting the temporal window available to generate the appropriate regulatory response (3,4). While these kinetic considerations were initially thought to be restricted to riboswitches regulating transcription, where an obligate decision must be made at the poly-uridine tract of the intrinsic terminator, it is becoming increasingly appreciated that translational riboswitches are also kinetically constrained (5–8). This is because these riboswitches also employ regulatory mechanisms such as rho-dependent transcriptional termination and RNA degradation in addition to occlusion or exposure of the ribosome binding site.

Rapid acquisition of a productive ligand-binding site within the aptamer domain is generally assisted by recurrent modules and structural motifs such as kink turns, T-loops, loop E motifs, and kissing loops to establish helical packing and long-range tertiary interactions (9–12). These motifs are typically identified as part of a bioinformatic analysis of a riboswitch's aptamer domain that yields a consensus secondary structure and nucleotide conservation pattern. For most riboswitch aptamer families, these modules are some of the most conserved sequence elements, underscoring their importance in folding and regulatory function (13,14). The consensus sequence for each of these recurrent modules is derived from a set of RNAs that vary sig-

*To whom correspondence should be addressed. Tel: +1 303 735 2159; Email: robert.batey@colorado.edu
Present address: Jacob T. Polaski, Computational Biology Program, Public Health Sciences Division. Basic Sciences Division, Fred Hutchinson Cancer Research Center, Seattle, WA 98109, USA.

nificantly in their overall structure, interaction with other RNAs or proteins, and biological function (15). However, each distinct family of RNAs likely place different demands on the structural module, such that within the consensus are subsets of sequences that are suitable for a given RNA's function. For example, the kink turn motif exhibits different preferences for sequences interacting with proteins versus those found in RNA-only structures (9). The kinetic constraints placed upon regulatory activity of riboswitches may favor a subset of variants within the consensus able to fold rapidly or form more stable long-range interactions in the absence of proteins. Differences in folding properties of the kink turn modules of ribosomes, riboswitches, and protein-binding snoRNAs have been observed (16–18). For modules found in highly represented riboswitches such as class-I cobalamin and S-adenosylmethionine-I (SAM-I), sequence alignment and comparative analysis with other families of RNAs readily identify variation in the module contributing to altered folding and functional properties that can be subsequently validated using biochemical approaches. For RNAs having a limited number of representatives or for modules that display substantial variation within a phylogenetic alignment, sequence analysis approaches are of limited utility.

Recently, a tertiary interaction mediated by a structure similar to the classic T-loop module (10) was found to be important for ligand binding in the class-II cobalamin riboswitch (19). Members of this class of cobalamin riboswitches, which contain two distinct T-loops (Figure 1A), exhibit a broad spectrum of selectivity for the two biological forms of cobalamin—5′-deoxyadenosylcobalamin (AdoCbl) and methylcobalamin (MeCbl)—such that some are highly selective for one of these forms while others show little discrimination between them (20). Furthermore, mutagenic analysis of an AdoCbl-selective (*env50*Cbl-IIa) and a MeCbl-selective member (*env8*Cbl-IIa) revealed that an internal loop element J1/3 (green nucleotides, Figure 1A), whose sequence is highly variable across a limited set of representative sequences, is a key selectivity determinant (see Supplementary Table S1). The last five nucleotides of the J1/3 internal loop (nucleotides 9–13 in the numbering of *env8*Cbl-IIa) form a turn structurally similar to a T-loop, and its long-range interactions with a region of the binding core J6/3 (magenta nucleotides, Figure 1A) is homologous to the T-loop/D-loop interaction of tRNA (Figure 1B). The T-loop motif is currently defined by a set of structure-based criteria that include: (i) five nucleotides that adopt a loop structure similar to a U-turn, (ii) a base pair between nucleotides 1 and 5 (A9 and U13 in Figure 1B), (iii) stacking of nucleotides 1 and 2, (iv) the nucleotide in position 4 either pairs with the nucleotide at position 2 or stacks on an intercalating base between nucleotides 4 and 5, and (v) the third nucleotide unstacked and free to form base-base interactions (10). Nucleotides 9–13 of J1/3 superimpose with a reference T-loop from *Escherichia coli* initiator tRNA with an RMSD of ∼0.55 Å, suggesting that J1/3 contains an authentic T-loop motif. However, this element might be more accurately classified as a lone pair triloop based upon bioinformatics analyses of this type of module found in other RNAs (21,22). The lone pair triloop motif is defined as a five nucleotide sequence containing a base pair between the first and fifth position and a three nucleotide hairpin loop between them. While the T-loop motif falls within this definition, the lone pair triloop encompasses a far broader range of closing pair identities and base stacking arrangements.

The lack of strong sequence consensus within J1/3 of class-II cobalamin riboswitches (Figure 1A) makes it hard to unambiguously classify this element using existing definitions. The main difference between the structure of J1/3 and the consensus T-loop motif is that the closing pair is a *trans* Watson–Crick base pair between nucleotides 9 and 13, where nucleotide 13 is in the *syn* orientation as dictated by the antiparallel arrangement of the two nucleotides, whereas the closing pair in the T-loop is most often a Watson–Crick/Hoogsteen U9–A13 pair (Figure 1C and D), although other pairings are accepted. The lone pair triloop sequence in the Gutell classification scheme that the J1/3 element of *env8*Cbl-IIa most closely conforms to is 'R1', which this scheme describes as having the closing pair in a reversed Hoogsteen conformation (Figure 2A) (21). Consideration of long range tertiary interactions with J6/3 would suggest that this module is better classified as an 'R2' triloop, and while J1/3 does contain a pyrimidine at position 3 of the triloop like R2 representatives, J1/3 does not conform to the requirement that the second triloop nucleotide is a uridine and makes hydrogen bonding contacts with the backbone (Figure 2B). In the Major classification scheme, which is primarily grouped according to how the bases stack, J1/3 is best described as a class I, 'SLSL' subclass triloop, of which T-loops are a member (Figure 2C) (22). While several members of this subclass have *trans* Watson-Crick/Watson-Crick closing pairs, their backbone orientation is classified as parallel, rather than the antiparallel arrangement of the closing pair in J1/3. Thus, classification of this important module in the class-II cobalamin riboswitches is difficult, and unfortunately, limited sequence conservation in J1/3 does not help to clarify this issue.

To gain further insights into the role of the J1/3–J6/3 interaction in supporting cobalamin binding and regulatory activity, we implemented a cell-based functional screen in which a set of critical nucleotides in each element were fully randomized and active variants identified. In this screen we identified 91 unique sequences with a broad range of abilities to repress expression of a fluorescent reporter protein, ranging from variants as efficient or moderately more active than the wild type riboswitch, to sequences that repressed reporter expression with only 10% of wild type activity. Analysis of these data reveal that RNAs with the highest repressive activity conform to a consensus in which the most functionally important elements are the three nucleotides in the J1/3 triloop and the Watson-Crick base pair between position 11 of J1/3 and position 70 of J6/3. This consensus also fits within T-loops found in other genetic elements. In addition, these results reveal that the J1/3–J6/3 interaction may not be essential for function, as a number of low activity variants contain no identifiable consensus in J1/3 but with an alternative consensus in J6/3. Thus, sequence alignments based upon this screen yield novel insights into both cobalamin riboswitch structure and function.

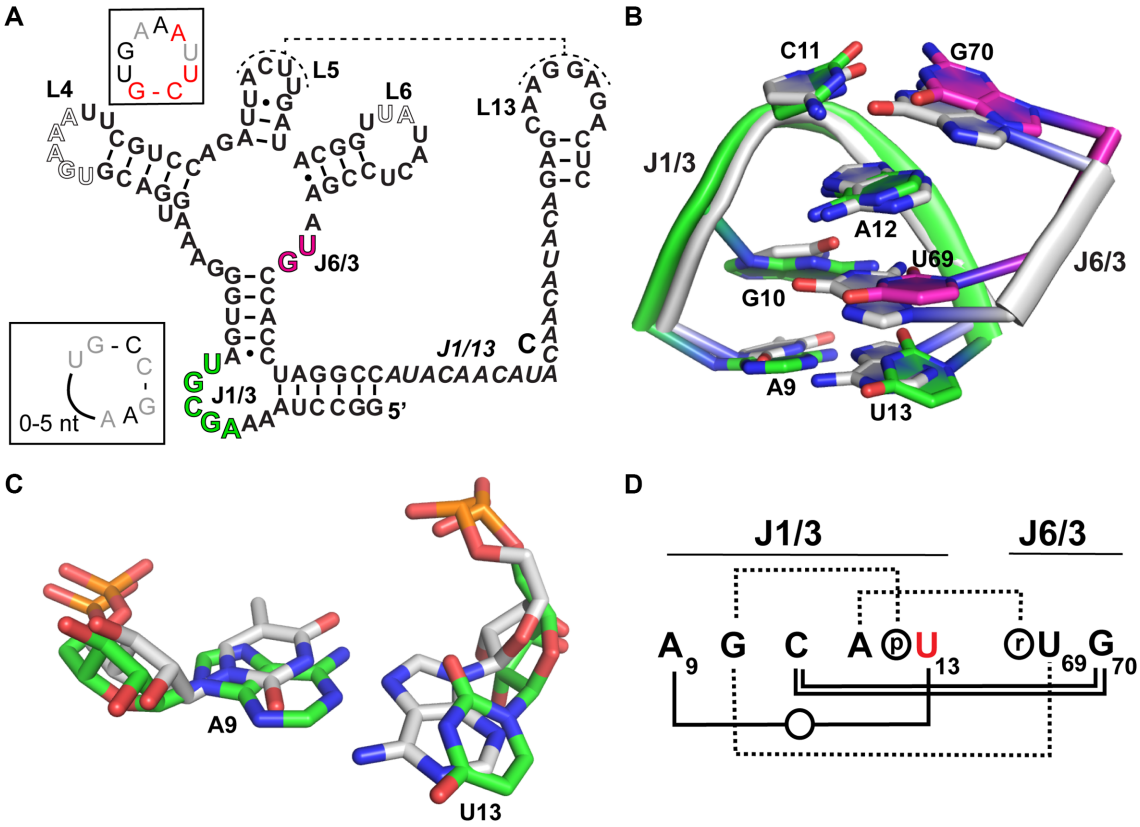

Figure 1. Structure of the *env8*Cbl-IIa riboswitch. (**A**) Secondary structure of the parental *env8*Cbl-IIa riboswitch with nucleotides that comprise the J1/3 internal loop colored green and nucleotides in J6/3 that interact with J1/3 colored magenta. Open letters denote nucleotides in L4 and L6 that are involved in a T-loop mediated loop-loop interaction. Nucleotides comprising the extended linker J1/13 are shown in black italics, and the regulatory loop-loop interaction between L5 and L13 that acts as the cobalamin-dependent regulatory switch is shown with a dashed black line. Insets show the consensus sequence of L4 and J1/3 from Weinberg *et al.* (50). Red nucleotides denote >97% conservation, black nucleotides denote >90% conservation and gray nucleotides >75% conservation. (**B**) Alignment of the canonical T-loop/D-loop interaction in *E. coli* initiator tRNA (Gray sticks; PDB: 3CW5) with J1/3 (green) and J6/3 (magenta) of the *env8*Cbl-IIa riboswitch (PDB: 4FRG; RMSD: 0.553 Å). Note that only two nucleotides of the D-loop that directly interact with the T-loop of tRNA are included. (**C**) Magnified view of the *env8*Cbl-IIa J1/3 closing Watson–Crick base pair between A9 and U13 (green sticks) and the closing Watson-Crick/Hoogsteen base pair in the tRNA T-loop (gray sticks). Note: The *env8*Cbl-IIa riboswitch was crystallized with an adenosine at position 12 of the J1/3 triloop rather than a guanosine that is present in the wild type sequence. (**D**) Schematic of base-base and base-backbone interactions within the J1/3–J6/3 interaction. The interaction nomenclature is that of Leontis and Westhof (51) in which bars or dashed lines represent base-mediated interactions, the circled 'p' represents a phosphate group, the circled 'r' represents a ribose sugar and the red nucleotide denotes a *syn* nucleotide.

MATERIALS AND METHODS

Library preparation and cloning

The riboswitch library was constructed using two single-stranded Ultramers® synthesized by Integrated DNA Technologies encompassing the entire riboswitch sequence with seven randomized positions (16 384 total variants), and NsiI and HindIII restriction sites at the 5′ and 3′ ends, respectively. Riboswitch inserts were amplified through 20 cycles of gap-filling PCR and were ligated into a reporter plasmid containing the fluorescent protein mNeon (23) and an ampicillin resistance cassette (Amp^R). Ligations were performed using the manufacturer's protocol for T4 DNA ligase (New England Biolabs) and 3 μl of the ligation reaction was transformed into 100 μl of chemically competent *E. coli* strain BW25113 (Δ*btuR*) by heat shock at 42°C for 30 s, which was recovered by adding 900 μl of 2x YT medium and incubated at 37°C for 1 h. 100 μl aliquots of the transformation reaction were plated onto chemically defined salt broth (CSB) media (24) containing 1.2% agar supplemented with cyanocobalamin (5 μM final concentration) and 100 μg/ml carbenicillin and grown overnight at 37°C. Once visible colonies formed, plates were incubated for an additional 24 h at 4°C to arrest cell growth and enhance the fluorescent signal. To generate counting statistics, colonies were counted following incubation at 4°C using the OpenCFU software package (3.9.0) (25).

Fluorescent colony screening

Colonies were illuminated using a light source outfitted with 490 nm excitation and 510 nm emission filters, respectively, for screening. Colonies exhibiting a similar level of fluorescence as the parent riboswitch (low on CNCbl plates) were picked and grown on gridded replica plates containing CSB media containing 1.2% agar and 100 μg/ml carbenicillin with and without 5 μM cyanocobalamin. Gridded plates were incubated at 37°C overnight to allow for cell growth

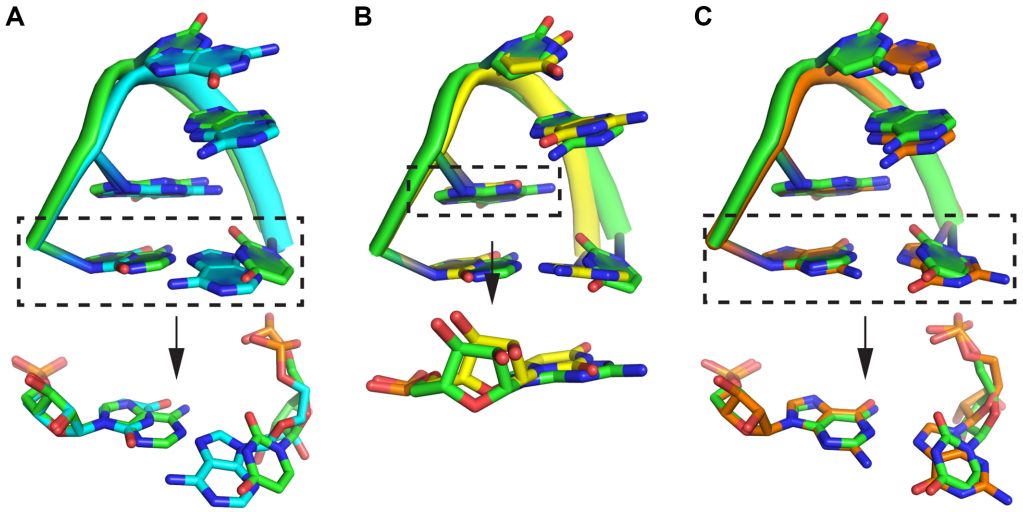

Figure 2. Superimposition of the *env8*Cbl-IIa riboswitch triloop motif with triloops found in other RNAs. (**A**) Superimposition of the *env8*Cbl-IIa triloop (green cartoon) with an R1-IA class triloop (cyan cartoon) from the Gutell classification found in the *H. marismortui* 50S ribosomal subunit (PDB: 1JJ2; RMSD: 0.198 Å). (**B**) Superimposition of the *env8*Cbl-IIa triloop (green cartoon) with an R2-IA class triloop (yellow cartoon) from the Gutell classification found in the *H. marismortui* 50S ribosomal subunit (PDB: 1JJ2; RMSD: 0.132 Å). (**C**) Superimposition of the *env8*Cbl-IIa triloop (green cartoon) with a class I SLSL triloop (orange cartoon) from the Lisi and Major classification found in the 16S rRNA of *T. thermophilus* (PDB: 1FJG; RMSD: 0.205 Å). For all superimpositions, the J1/3 element of *env8*Cbl-IIa is in the identical orientation as depicted in Figure 1B.

and were subsequently incubated at 4°C for 24 h to arrest cell growth and enhance the fluorescent signal.

Colonies showing cobalamin-dependent regulation of mNeon upon visual inspection were picked, transferred to 10 ml of 2xYT media supplemented with 100 μg/ml ampicillin, and grown overnight to saturation at 37°C. Plasmids were isolated from overnight cultures via miniprep (E.Z.N.A.® Plasmid Mini Kit, OMEGA bio-tek) the following day and analyzed by Sanger sequencing. Plasmids harboring unique riboswitch sequences were transformed into chemically competent *E. coli* BW25113 (Δ*btuR*) cells, plated onto CSB medium containing 1.2% agarose supplemented with 100 μg/ml carbenicillin and grown overnight at 37°C.

Riboswitch activity assays

To determine levels of mNeon expression, three colonies for each riboswitch were picked and grown overnight to saturation in 5 ml of CSB media supplemented with 100 μg/ml ampicillin. The following day, 5 μl of each overnight culture was transferred to tubes containing 5 ml of CSB medium with and without 5 μM cyanocobalamin supplemented with 100 μg/ml ampicillin. Cultures were grown to mid-log phase at 37°C in a roller drum, at which time 300 μl of each culture was added to wells in a Costar® 96-well half area microplate. Expression of mNeon was measured at an excitation wavelength of 490 nm and a 517 nm emission wavelength using a Tecan Infinite M200® PRO plate reader.

Levels of mNeon expression were determined from a single technical replicate of three biological replicates, which were normalized to the cell density (OD_{600}) in each well and background corrected using cell density normalized fluorescence from wells containing a pBR327 empty vector. Fold repression was calculated by dividing the average normalized background corrected fluorescence of the unrepressed condition (absence of cyanocobalamin) by the average normalized background corrected fluorescence of the repressed condition (presence of cyanocobalamin).

Data processing and phylogenetic tree construction

Sequence alignments were performed using the Multiple Sequence Comparison by Log-Expectation (MUSCLE) alignment tool in Jalview and the phylogenetic tree was generated using the Geneious software package (version 11.0.2, https://www.geneious.com) (26). Tree distances and sequence relatedness was determined by the method of Tamura and Nei (27) using a neighbor-joining model with no outgroup, which was converted to a sunburst format to aid visualization.

RESULTS

Design and implementation of the screen

To obtain a functionally active library of J1/3–J6/3 variants of the *env8*Cbl-IIa riboswitch, nucleotides 9–13 of J1/3 and nucleotides 69 and 70 of J6/3 were fully randomized to yield 16 384 theoretically possible variants (red dots, Figure 3). This complete variant library was cloned into a reporter plasmid with the riboswitch in the leader sequence of an mRNA encoding the fluorescent protein mNeon (23), whose transcription is driven by a moderately strong synthetic insulated constitutive promoter 'proD' (28). This reporter system has been previously validated and used to investigate the mechanism of cobalamin-dependent gene regulation and ligand selectivity of the *env8*Cbl-IIa cobalamin riboswitch (5,20). To increase the dynamic range of the regulatory response, an *env8*Cbl-IIa variant was used that features an extension of a linker sequence (J1/13) connecting the receptor and regulatory domains from 7 nucleotides in

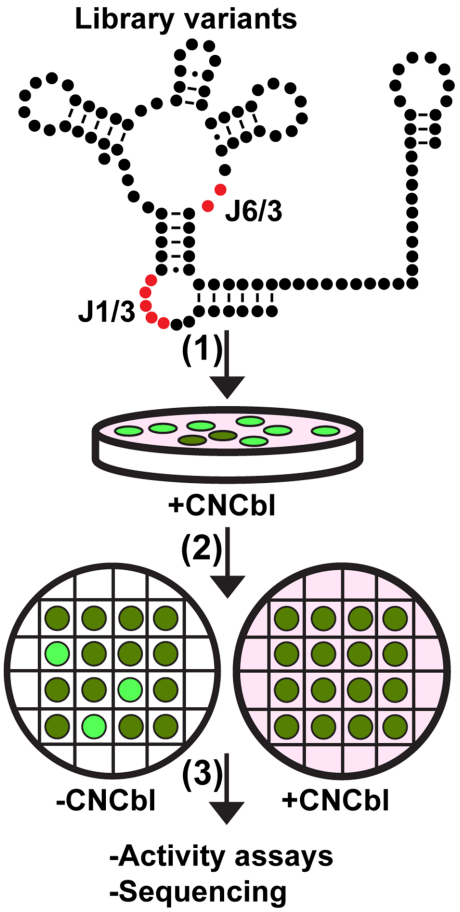

Figure 3. General workflow for the cell-based screening strategy. A simplified schematic of the parental *env8*Cbl-IIa riboswitch is featured at the top with the randomized nucleotides in J1/3 and J6/3 colored red. In step (1), library variants are transformed into *E. coli* and plated on CSB agar containing CNCbl. In step (2), colonies are selected that exhibit similar levels of mNeon expression as the parental sequence and transferred to gridded plates where they are grown either in the presence or absence of CNCbl. In step (3), colonies that exhibit CNCbl-dependent repression of mNeon expression upon visual inspection are harvested and used for activity assays in liquid culture and sequenced.

the wild type riboswitch to 19 nucleotides (5). Lengthening this region of the RNA increases regulatory activity by increasing the amount of fluorescent protein expression in the absence of cobalamin while retaining the ability to repress gene expression to wild type levels in the presence of ligand. This alteration does not affect ligand binding to the aptamer but does improve the ability to visually sort riboswitch variants that demonstrate cobalamin-dependent gene regulation. The library of plasmids was transformed into *E. coli* strain BW25113 (*ΔbtuR*), which is defective in cob(I)yrinic acid a,c-diamide adenosyltransferase, an enzyme that adenosylates cobalamins during import (29,30), ensuring that cyanocobalamin (CNCbl) used in the growth medium is not converted to AdoCbl.

To ensure that colonies were sufficiently separated for visual screening while maximizing the degree of library coverage per round of screening, transformation reactions were optimized to yield ∼150–200 colonies per 60 mm diameter plate. To obtain CNCbl-responsive riboswitches, the transformed library was screened using a two-step approach. In the first step, the transformed library was plated on a rich, chemically defined medium (24) containing 5 μM CNCbl (step (1)), Figure 3). Colonies were visually screened for expression of the mNeon reporter and manually picked if they exhibited a fluorescent signal similar to the parent riboswitch (dark colonies, Supplementary Figure S1). Selected colonies were then grown on gridded plates (step (2), Figure 3) in the absence or presence of 5 μM CNCbl and cells that showed visual differences in their brightness were used for downstream analysis and characterization (step (3), Figure 3).

In four rounds of screening, 74 300 colonies were surveyed, representing ∼4.5-fold redundancy (Supplementary Table S2). Theoretical calculation suggests this level of redundancy results in ∼99% probability of observing each unique sequence in the original pool of variants (31), but in practice, this is likely lower. In total, four rounds of screening yielded ∼2578 colonies that were selected for secondary analysis using gridded plates, which resulted in 146 variants that exhibited visually detectable CNCbl-dependent repression of mNeon. Each positive switch was sequenced, resulting in 91 unique variants (Supplementary Table S3). It should be noted that in the final round of screening, we still observed 24 unique sequences from 40 possible variants from the grid screen, indicating that screening with 4.5-fold redundancy was not sufficient to fully capture all active sequence variants. However, this number of unique sequences is adequate to provide a clear alignment in which conservation patterns would emerge.

Riboswitch variants cluster into well-defined classes based on sequence

Analysis of the functionally active variant library was initiated by sequence alignment of the 91 variants using the Multiple Sequence Comparison by Log Expectation (MUSCLE) software package (Supplementary Figure S2) (32), and a cladogram calculated to visualize related sequence clusters using the neighbor joining method of Tamura and Nei (Figure 4A) (27). At the topmost level, the cladogram is readily divided into three major classes (yellow, cyan and magenta sectors) with two outlying sequences that do not group with the three major clusters (green sector). The cladogram can be further divided into eight subclasses (SC-I to SC-VIII) to provide more detailed information about conservation patterns within smaller related sub-clusters. Additionally, analysis of consensus sequences was informed using activity assays performed in liquid culture to determine the fold repression of all 91 unique variants (red circles, Figure 4A and Supplementary Figure S3 and Supplementary Table S4).

Highly active variants cluster into one major grouping

One major class of the cladogram is largely populated with sequences exhibiting strong repressive activity and includes the parental riboswitch (red star, Figure 4A). The consensus sequences of J1/3 and J6/3 show that members of this class generally preserve the Y11–G70 base pair observed in the parental riboswitch, have a nearly invariant guanosine at position 10, a near equal preference for an adenosine

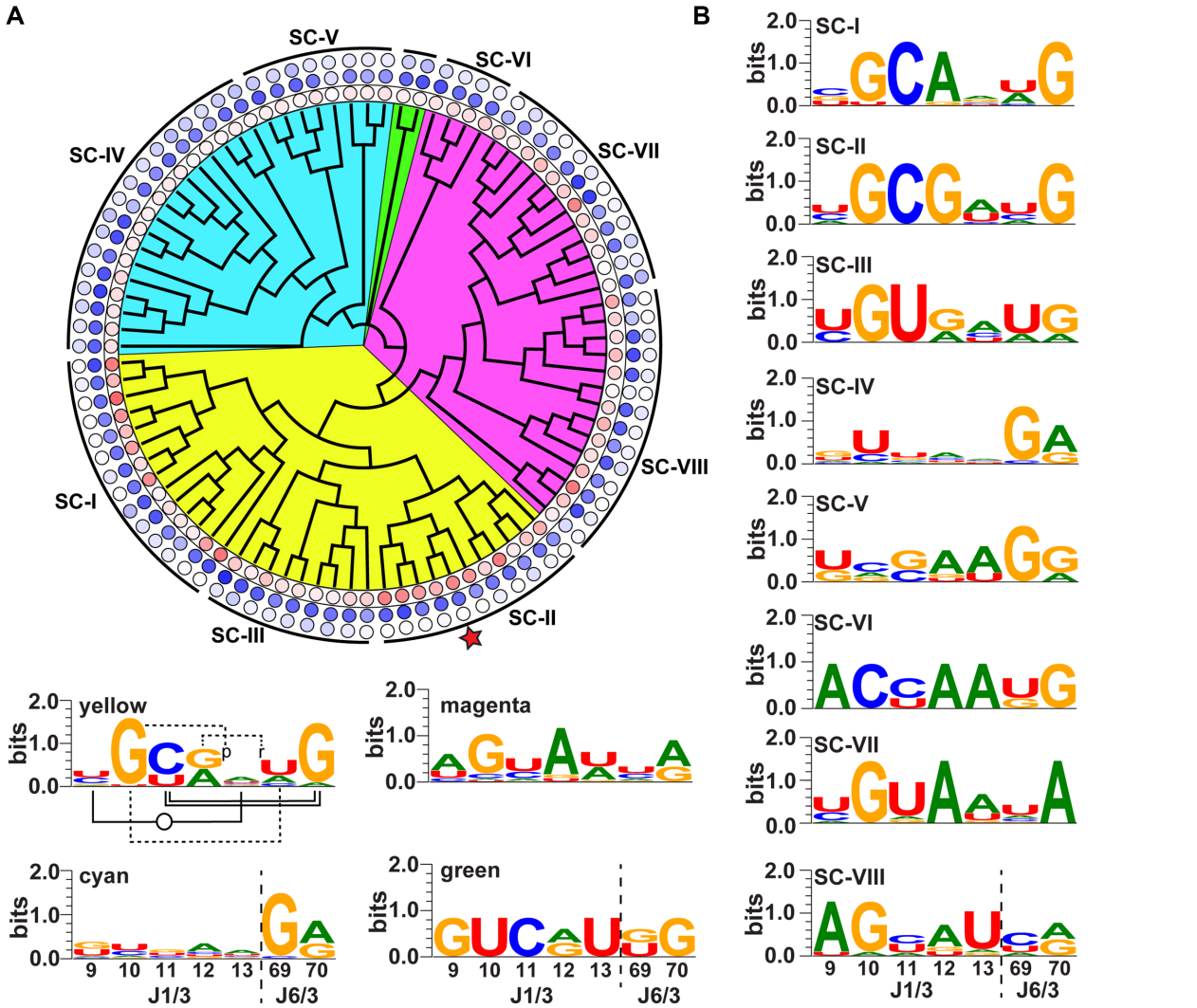

Figure 4. Cladogram and consensus sequences for variants found in the screen. (**A**) Sunburst style cladogram depicting well-defined clusters of related sequences along with activity assay data obtained from liquid culture assays. Major sequence classes are separated using colors (yellow, cyan, and magenta), with two unrelated outlying sequences highlighted in green. Relative fold repression values from activity assays for each variant is represented with a red circle in the first shell, with darker red circles denoting riboswitches with higher dynamic ranges. Blue circles in the second shell represent relative values of cell density normalized fluorescence for cells grown in the absence of CNCbl, and blue circles in the outermost shell represent cell density normalized fluorescence of cells grown in the presence of CNCbl. In both cases, circles with a darker blue color indicate higher levels of mNeon expression. The three major classes are further divided into subclasses, which are shown using the designation 'SC-X', and the parental riboswitch used to design the library is denoted with a red star in the yellow sector. (**B**) Sequence logo representation (52) of consensus sequences for J1/3 (nucleotides 9–13) and J6/3 (nucleotides 69 and 70) for the subclasses shown in panel *A*. All sequence logos were generated using the WebLogo 3.5.0 software package (53). The secondary structural interactions in wild type J1/3–J6/3 are shown in the 'yellow' panel and use the same notation as Figure 1D. Note that these interactions may not be present in all sequences.

or guanosine at position 12, and a moderate preference for a uridine at position 69 as the intercalating base. Notably, members of this class show a high diversity of nucleotides comprising the closing base pair between positions 9 and 13 of J1/3, including non-Watson–Crick C–A, C–C, U–U, C–U and A–A pairs in sequences with high dynamic range, indicating this portion of the loop is highly tolerant of diverse potential base pairs.

Within this class, a subset of sequences (SC-I, Figure 4B) has a consensus in which three nucleotides of J1/3 (nucleotides 10–12) and position 70 of J6/3 are almost invariant, suggesting the most important feature of the tertiary interaction is maintaining a Watson–Crick C11–G70 base pair, which is invariant in the group. The structural role of nucleotide 10 is to stabilize the turn in J1/3 via interaction between the Watson-Crick face of guanosine and the phosphate moiety of nucleotide 13. It should be noted that within the T-loop and the lone triloop module, the nucleotide corresponding to nucleotide 10 is often a uridine, whose base can make an equivalent set of contacts with the backbone (10). Like the yellow class as a whole, the closing pair of J1/3 shows no conservation pattern or pairing preference–potential C–C, C–U, U–A, C–A, G–G and U–G pairs are observed, which can all potentially form a reverse Watson–Crick or a wobble base pair with two hydrogen bonds. However, the wild type A9–U13 pair was not ob-

served, which likely reflects that the screen did not capture all possible active variants. Position 69 shows a near equal preference for uridine or adenosine, with only one member containing a cytosine at this position; it is not clear why cytosine and guanosine should be so strongly selected against. The variant with the lowest dynamic range in this subclass (GGCGG/UG, 2.2-fold repression) is the only member that has a guanosine rather than an adenosine at position 12. However, it should be noted that in all members of SC-II discussed below, which is also enriched in highly active variants, this position is always a guanosine, indicating this alone does not abrogate regulatory activity.

A second subset of sequences (SC-II, Figure 4B) is largely comprised of highly active variants, and not surprisingly, the consensus within this subclass is very similar to SC-I. What differentiates this class from SC-I is the nucleotide identity of position 12, which is exclusively a guanosine rather than an adenosine. Within this subclass, the least active variant (CGCGC/CG, 2.9-fold repression) is the only member with a potential C9–C13 pair in J1/3, a type that is represented in SC-I in two highly active sequences (CGCAC/UG, 14-fold repression and CGCAC/AG, 12-fold repression), suggestive of covariation between the identity of position 12 and the closing base pair in J1/3. Overall, SC-I and SC-II contain sequences most likely to exhibit robust gene regulation in a biological context, a result consistent with the presence of the parental sequence in SC-II (star, Figure 4A).

The third subclass of the yellow sector (SC-III, Figure 4B) contains variants with modestly lower activity (2.8–7.2-fold repression) than sequences in SC-I and SC-II. The major difference between SC-III compared to SC-I and SC-II is the presence of a U11–G70 base pair in all sequences except for one with low dynamic range (UGUGA/AA, 3.0-fold repression). SC-III is also enriched in sequences that contain either a U9–A13 or U9–U13 closing base pair in J1/3, which is found in more highly active SC-I representatives, suggesting that a more optimal closing lone pair is required to compensate for a weaker 11–70 G–U pair. Additionally, SC-III has a strong preference for an intercalating uridine at position 69; one sequence has an adenosine at this position and represents the lowest activity member (UGUGA/AG, 2.8-fold repression). An adenosine at position 69 in SC-I and SC-II is well-tolerated, suggesting that the presence of the U11–G70 wobble pair imposes a requirement for the more optimal uridine at position 69. Together, the sequences of SC-III reveal that deviations from the optimal triloop and Watson–Crick 11–70 pair can be compensated for by other features of this module.

A second major class largely includes sequences with moderate activity

The second major cluster of variants in the cladogram is populated by sequences exhibiting moderate dynamic range compared to the parental riboswitch (magenta sector, 5.4 median fold repression; Figure 4A). Like the parental sequence, this class contains many variants (11/30 sequences) that have the potential to form a closing Watson–Crick/Hoogsteen base pair between A9 and U13 in J1/3, as well as many variants (14/30 sequences) with a uridine as the intercalating base at position 69 and an adenosine at position 12 (26/30 sequences), which is also observed in members of the yellow class. However, unlike the parental riboswitch and RNAs in the yellow class, only five variants in the magenta class can form a C11–G70 Watson–Crick pair between J1/3 and J6/3 (most variants have a U11–A70 base pair at this position), further reinforcing that this tertiary contact plays a significant role in the assembly of the cobalamin-binding core and efficient gene regulation by the RNA. The loss of the optimal C11-G70 pair is likely to be partially compensated for by optimal closing 9–13 pairing.

Within this major class, a subset of variants (SC-VI, Figure 4B) with moderate dynamic ranges (3.9- to 5.7-fold repression) feature a conserved non-Watson-Crick closing A-A base pair in J1/3, an invariant adenosine at position 12, and either a C11–G70 or U11–G70 base pair between J1/3 and J6/3. Notably, although members of this subclass contain sequence features present in more active RNAs from SC-I and SC-II, the main difference in SC-VI representatives is a nearly invariant cytosine at position 10 (3/4 sequences), which is almost always a guanosine in SC-I and SC-II. While the nucleotide identity of position 10 could have a strong effect on regulatory function, the low number of sequences in this subclass (4) makes it difficult to draw strong conclusions using its consensus.

A second subclass within the magenta class (SC-VII; Figure 4B) contains variants with a broad spectrum of dynamic ranges (1.2- to 14.0-fold repression), but is largely populated with sequences exhibiting regulatory activity near the center of this distribution (5.3 median fold repression). SC-VII features three highly conserved positions within J1/3 (nucleotides G10, U11 and A12) and an invariant adenosine at position 70 in J6/3. This means that, unlike members of highly active subclasses SC-I and SC-II, SC-VII representatives almost invariably have a U11–A70 base pair rather than a C11–G70 pair. Notably, the variant of SC-VII with the highest dynamic range (CGUAA/UA; 14.0-fold repression) features a suboptimal U11–A70 base pair, which may be offset by the presence of G10 and an intercalating uridine that is found in highly active SC-I and SC-II representatives. Members of SC-VII also show some preference for a U–A as the closing pair of the J1/3 triloop, however this preference is not strong, as RNAs with A–A, U–U, C–A and U–G pairs are also found in this subclass.

Like SC-VII, SC-VIII is populated by variants with a broad range of regulatory activities (1.6- to 15.2-fold repression), but on average contains sequences with similar dynamic ranges as SC-VII (5.5 median fold repression). SC-VIII representatives show a strong preference for a closing A–U pair in the J1/3 triloop and have a lower degree of nucleotide conservation at positions 11 and 70; variants with U–A, C–A, U–G, A–G and C–G base pairs are found in this subclass. Similar to highly active RNAs found in SC-I and SC-II, the consensus for SC-VIII shows an almost universal conservation of a guanosine at position 10, but unlike all other subclasses shows a moderate preference (8/14 sequences) for a cytosine as the intercalating base at position 69.

A third major class contains variants with low levels of regulatory activity

The third major class (cyan sector; Figure 4A) is dominated by sequences with low repressive activity compared to the parental sequence (the highest is 5.0-fold repression with sequence CUUAA/GA) and significantly less nucleotide conservation throughout J1/3 compared to other subclasses. The most highly conserved nucleotide in this class is the intercalating base at position 69, which is predominantly a guanosine.

There are two subclasses within this class (SC-IV and SC-V; Figure 4B) that show low degrees of nucleotide conservation in J1/3 and J6/3 compared to all other subclasses. Like the cyan class as a whole, variants comprising SC-IV show nearly universal conservation of a guanosine at position 69, a moderate preference for a uridine at position 10, and either adenosine or guanosine at position 70. While representatives of the second subclass (SC-V) show universal conservation of a guanosine at position 69, these RNAs show a moderate preference for a U9–A13 closing pair of the J1/3 triloop, which is the most highly represented base pair (4/8 sequences). Like highly active variants in SC-I and SC-II, some riboswitches in SC-V feature a C11-G70 base pair (3/8 sequences) between J1/3 and J6/3; however, other sequences in this subclass have non-Watson–Crick G–A or G–G pairs at this position. This suggests that while some members of the cyan class contain the optimal C11–G70 base pair, a combination of other non-optimal features in the motif contribute to the lower dynamic ranges observed in both subclasses.

The two outlying sequences (green sector; Figure 4A) have J1/3 and J6/3 sequences: GUCAU/GG and GUCGU/UG and exhibit low dynamic range compared to the parental sequences (3- and 2-fold repression, respectively). Although both sequences can form the optimal C11–G70 base pair observed in active variants comprising SC-I and SC-II, other sub-optimal features, including a guanosine at position 9, a uridine at position 10, and for one sequence a guanosine as the intercalating base, likely diminish the regulatory activity of these two variants.

Analysis of consensus sequences by activity-based clustering

An alternative way to group sequences obtained from the screen is by their dynamic range relative to the wild type sequence (14.0-fold repression of reporter gene expression). To achieve this, variants were categorized into four groups: >75%, >50–75%, >25–50% and >10–25% of wild type activity, as described below.

Riboswitches with >75% wild type activity

This group contains all sequences exhibiting greater than or equal to 10.5-fold repression of mNeon expression in the presence of CNCbl. The consensus sequences of J1/3 and J6/3 in this cluster (Figure 5A) are highly similar to SC-I and SC-II, which is expected as the majority of these sequences originate from these subclasses. The only members not from SC-I and SC-II are AGCAU/CG (15.2-fold repression) and CGUAA/UA (14.0-fold repression), found within SC-VIII, and SC-VII, respectively. Overall, the consensus indicates a strong preference for guanosine and cytosine at positions 10 and 11 in the J1/3 triloop, respectively, and a modest preference for either adenosine or guanosine at position 12. Additionally, like the parental sequence, members of this cluster show a preference for a uridine as the intercalating base at position 69; however, 4/15 RNAs tolerate an adenosine in this position and 2/15 feature a guanosine. Members of this cluster also have an invariant guanosine at position 10, which forms key contacts with the backbone to enforce the turn of the triloop, as well as a strong preference for a C11–G70 Watson–Crick base pair between J1/3 and J6/3 (only 1/15 sequences has an alternative U11–A70 base pair). The consensus for J1/3 shows a weak preference for the identity of the closing base pair between positions 9 and 13, which is noteworthy because if all the sequences in this cluster conform to having a reverse Watson-Crick or wobble (9-*anti*/13-*syn*) base pair scheme, seven types of pairs can support this arrangement (C–A, U–A, A–U, C–C, U–U, A–A and U–G).

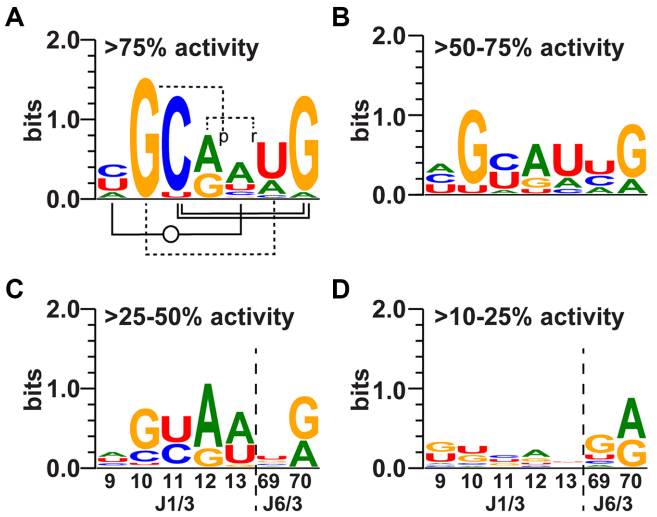

Figure 5. Sequence logos of J1/3 (nucleotides 9–13) and J6/3 (nucleotides 69 and 70) of riboswitch variants grouped according to activity relative to the parental construct. (**A**) Consensus sequences of riboswitches that achieve 75–100% activity of the parental RNA. The secondary structural interactions in wild type J1/3–J6/3 are shown and use the same notation as Figure 1D. (**B**) Consensus sequences of riboswitches that exhibit 50–75% activity of the parental RNA. (**C**) Consensus sequences of riboswitches with 25–50% activity of the parental RNA. (**D**) Consensus sequences of riboswitches with 0–25% activity of the parental RNA.

Riboswitches with >50% to 75% wild type activity

The consensus sequences of J1/3 and J6/3 in this group are substantially relaxed compared to the most active RNAs (Figure 5B). In this cluster, G10 remains the most highly conserved position, although uridine is also infrequently observed at this position (2/9 sequences). The identity of the intercalating base at position 69 is more variable than the more active sequence cluster, and positions 11 and 70 show a lesser preference for a C11–G70 Watson–Crick pair, with a U11–A70 base pair occurring in 2/9 and a U11–G70 pair in 2/9 representatives. There remains a strong prefer-

ence for a purine at position 12, although there is a single example of a uridine at this position, which is also the sequence that contains a non-Watson–Crick A11–G70 base pair (UGAUU/CG). Finally, the identity of nucleotides comprising the closing pair between positions 9 and 13 are similar to the most active group, with the exception of a single sequence with a C9–U13 pair that is also capable of forming a reverse Watson–Crick pair with two hydrogen bonds. This arrangement would place the C(O2) and U(O4) in close proximity, suggesting that an unstable pair or even no pairing at this position still supports the tertiary interaction between J1/3 and J6/3.

Riboswitches with >25% to 50% of wild type activity

This large group of variants (34 sequences) has a further relaxed consensus sequence compared to the two more active clusters (Figure 5C). The preference for a guanosine at position 10 is significantly reduced, particularly among members with the lowest repressive activity, which typically have a cytosine or adenosine. Within this group there appears to be no preference for a particular nucleotide at position 69, suggesting that the weak interaction between nucleotides 10 and 69 is not critical for establishing the interaction between J1/3 and J6/3. Furthermore, while many members have a Watson–Crick base pair between positions 11 and 70, potential non-Watson–Crick pairs at this position are increasingly observed, including U11–G70 (6/34 sequences), C11–A70 (2/34 sequences), G11–G70 (1/34 sequences) and A11–G70 (1/34 sequences). However, despite the increased variation in pairing schemes between positions 11 and 70, the strong preference for Y11–R70 remains. Similar to the more active clusters, there is only weak conservation of the identity of the closing base pair between positions 9 and 13, with G9–G13 and G9–A13 pairs observed along with the previously noted base pairs.

Riboswitches with >10% to 25% of wild type activity

This group represents a set of 32 RNAs that exhibit weak regulatory activity that ranges from 3.4- to 1.6-fold repression. The consensus of the J1/3 element of this cluster does not show a conservation pattern consistent with the lone pair triloop motif or T-loop module (Figure 5D). Instead, the only well-defined consensus occurs in J6/3 at position 70, which is always a purine, but less than half (14/32) have a nucleotide at position 11 that can form a Watson–Crick base pair. Representatives of this cluster also show a modest preference (19/32 sequences) for a guanosine as the intercalating base at position 69, which deviates from the preference for a uridine in more highly active sequences. Overall, the consensus sequences of members from this group suggests the interaction between J1/3 and J6/3 in these RNAs is weak or potentially absent. In this case, the lack of this tertiary interaction weakens the affinity of the binding pocket for cobalamin (19), which impairs the ability of the RNA to inhibit translation initiation.

DISCUSSION

In this work, the sequence preferences of a key tertiary interaction in a class-II cobalamin riboswitch was investigated using a cell-based functional screen to generate an experimental phylogeny of an RNA structural module critical for cobalamin selectivity and regulatory function. This approach allows for the unbiased generation of a diverse set of unique riboswitch variants, where differences in the sequence composition and their impact on regulatory activity are assessed in the same overall structural context. A screening-based approach with manual colony screening was chosen because this ensured we would be able to access variants that span a broad spectrum of repressive activities, rather than only the most active variants that would likely emerge from selection approaches. These low activity variants have the potential to reveal aspects of the module or RNA of interest that the most active sequences cannot. Colony screening also enables functional testing of each active variant in an integrated workflow to yield a dataset that reveals correlations between sequence and regulatory activity. While this result can be achieved using higher-throughput approaches, the limited size of this library and nature of the activity screen made colony screening an optimal strategy despite being labor-intensive. Furthermore, this approach is generalizable to study myriad RNA structural motifs, which can be used to understand evolutionary trajectories and fitness landscapes of a wide range of functional RNAs, as has been done previously for proteins (33,34) and catalytic RNA (35).

The screening strategy employed in this study yielded 91 unique riboswitch variants with a broad range of biological activity. The most active variants cluster into three subclasses (SC-I, SC-II and SC-III) that together share a highly similar consensus. These RNAs do not have a strong preference for the identity of the closing pair of the lone pair triloop in J1/3, which is consistent with previous large-scale surveys of this motif in various rRNA and tRNA crystal structures that revealed the presence of diverse non-Watson-Crick base pairs at this position (21). Instead, the most important features are the sequence of the triloop itself and the Watson-Crick pair between nucleotides 11 and 70. SC-I and SC-II representatives have a universally conserved C11–G70 Watson–Crick base pair between J1/3 and J6/3, while the consensus of members from SC-III, which are generally less active than SC-I and SC-II variants, feature either U11–G70 or U11–A70 base pairs at this position. Overall, the nucleotide conservation patterns of sequences supporting >50% of wild type activity strongly correlate with the T-loop conservation pattern in genetic elements (Figure 6) (10). Thus, this element should be classified as the second T-loop motif in the Cbl-II riboswitch family—the first being L4 (Figure 1A). It should be noted that the T-loop is part of the larger lonepair triloop motif as defined by Gutell (21) and Lisi and Major (22).

The crystal structure of the *env8*Cbl-IIa aptamer in complex with hydroxocobalamin provides structural insights into the observed sequence preferences (19). A strong base pairing interaction between J1/3 and J6/3 likely serves to prevent alternative pairing between J3/4 and J6/3 that would collapse the cobalamin binding pocket, thereby preventing translational repression. A rapidly acquired and stable J1/3 structure is thus likely to be important in establishing the cobalamin binding pocket on a timescale conducive to regulatory function. Our data suggest that a

A Nucleotide Frequences, >50% wild type activity J1/3 T-loop

Position	1	2	3	4	5
	A = 24%	A = 0%	A = 4%	A = 64%	A = 48%
	G = 0%	G = 96%	G = 0%	G = 32%	G = 4%
	C = 36%	C = 0%	C = 72%	C = 0%	C = 12%
	U = 50%	U = 4%	U = 24%	U = 4%	U = 36%

B Nucleotide Frequences, genetic elements (introns and riboswitches)

Position	1	2	3	4	5
	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 = 6.1%	C = 0%	C = 27.3%	C = 0%	C = 12%
	U = 81.8%	U = 9.1%	U = 0%	U = 0%	U = 9.1%

Figure 6. (**A**) Nucleotide frequencies for each of the five positions in J1/3 for riboswitch variants discovered in the screen with >50% wild type activity. (**B**) Nucleotide frequencies for each of the five positions in T-loops identified in naturally occurring introns and riboswitches for which there is structural information. Data is taken from reference (10).

guanosine at position 10 is particularly critical for this. Interestingly, this position is weakly conserved across triloops (22), but within T-loops there is a strong preference for uridine at this position in tRNAs and guanosine in rRNA and other RNAs (10). Consistent with this hypothesis, our data generally show similar levels of mNeon expression across all variants in the absence of cobalamin, but sequences with lower repressive activity typically allow for higher levels of gene expression in the presence of ligand, reflecting a lower propensity to bind cobalamin (Figure 4A).

At the lowest end of the activity spectrum, J1/3 no longer has any pattern of conservation. In contrast, J6/3 still displays a weak preference for guanosine or uridine at position 69 and a strong preference for adenosine or guanosine at position 70. This suggests that at least some of the weakly active variants may lack the tertiary interaction between J1/3 and J6/3 but still bind cobalamin in some capacity to regulate gene expression. It should be noted that there are a number of members within the Cbl-II family—a well characterized example was discovered in the gut commensal bacteria *Enterococcus faecalis* (36)—that lack an identifiable J1/3 element and may function without a tertiary interaction that helps to organize the cobalamin-binding core.

These data reveal new insights into the nature of the lone pair triloop motif and T-loop module that have broad implications for being able to accurately model RNA structure. One approach to computational modeling of RNA 3D structure is knowledge-based fragment assembly in which a crude model is assembled from a set of modules and then refined to yield a final model (37). This avenue has worked well for being able to predict and model RNA structures with increasing accuracy (38,39). In part, this approach relies upon the observation that highly recurrent RNA modules such as the kink turn and the GNRA tetraloop share a common structure and are significantly context independent (40,41). However, this does not appear to be the case for the lone pair triloop motif, of which the T-loop is part. Many of the sequences that are observed in the most efficient riboswitches fall into different classes and subclasses of triloops in the classification scheme developed by Lisi and Major (22). We find diverse J1/3 sequences within multiple subclasses of Class I and IV triloops as well as in Class III, VI and VII. Since each of these classes are defined by the structure of the triloop and associated lone pair, this would predict that these riboswitches adopt different structures of J1/3, some of which would be inconsistent with the crystallographically observed interaction with J6/3 (19). Instead, it is most likely that all sequences adopt a T-loop (Class-I, subclass SLSL* in the Lisi and Major classification scheme). This would indicate that the structure of a sequence that is part of a lonepair triloop is highly context dependent, making this motif potentially difficult to accurately model based upon a particular classification.

These data support that sequence preferences of the T-loop module are dependent on both its context and the biological function of the RNA. With regard to context, T-loops in other genetic elements—introns and riboswitches—are deplete of uridine at position 3. But in J1/3, uridine is significantly represented, reflecting the strong preference for a Y11–R70 base pair. Furthermore, cytosine, which is not tolerated across all structurally defined T-loops (10), is observed in our data. In fact, the strong preference for a *trans* Watson–Crick-Hoogsteen U–A pair in the T-loop motif is significantly less in J1/3, suggesting that this T-loop is more accommodating of diversity in the closing pair. This likely reflects the functional role of the T-loop in the riboswitch. Another constraint for the sequence of the J1/3 T-loop is the kinetic pressure to fold both rapidly and reliably, as the coupling of transcription and translation in bacteria (42,43) places strong temporal requirements on sensing the occupancy status of the receptor to produce the appropriate regulatory outcome (44,45). Furthermore, many RNAs fold under the influence of proteins and chaperonins (46,47), which places a different set of evolutionary constraints on their sequence composition, as the RNA bases must make productive contacts with amino acids. Thus, it is not surprising that there are clear differences in sequence preferences for T-loops found in tRNAs, rRNAs and between different genetic elements (10).

Ultimately, for riboswitches and other functional RNAs, the cellular environment places constraints on the allowable sequence space to achieve proper folding and the desired cellular function (48,49). The screening strategy in this work uncovered dozens of variants with dynamic ranges spanning a broad range of repressive activity, suggesting the sequence composition of the J1/3-J6/3 interaction plays an important role in the folding of this riboswitch and the subsequent regulatory response. While additional work is required to determine the effect of different J1/3 T-loop sequences on folding kinetics, in the case of cobalamin riboswitches, this element likely serves a central role in rapidly establishing a productive binding pocket for cobalamin to ensure repression of gene expression on the timescales required for a co-transcriptional/translational regulatory mechanism. Recurrent motifs in many other functional RNAs may act in a similar fashion as critical nodes along folding pathways, and developing a more comprehensive understanding of allowable and optimal sequence space for these modules will have broad implications for RNA structure prediction and the design of synthetic RNA-based devices.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank Prof. Amy E. Palmer for granting us access to their light source for fluorescence-based colony screening. *Author Contributions*: R.T.B. and J.T.P. conceived of the project with assistance from L.K.D. about the design, implementation, and optimization of the functional screening strategy. J.T.P. performed colony screening and characterization of individual riboswitch variants with assistance from O.A.K. R.T.B. and J.T.P. prepared the manuscript with input from all authors.

FUNDING

National Institutes of Health (NIH) [R01 GM073850 to R.T.B.]. Funding for open access charge: NIH.
Conflict of interest statement. None declared.

REFERENCES

1. Garst,A.D., Edwards,A.L. and Batey,R.T. (2011) Riboswitches: structures and mechanisms. *Cold Spring Harb. Perspect. Biol.*, **3**, a003533.
2. Roth,A. and Breaker,R.R. (2009) The structural and functional diversity of metabolite-binding riboswitches. *Annu. Rev. Biochem.*, **78**, 305–334.
3. Lemay,J.F., Desnoyers,G., Blouin,S., Heppell,B., Bastet,L., St-Pierre,P., Masse,E. and Lafontaine,D.A. (2011) Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLos Genet.*, **7**, e1001278.
4. Garst,A.D. and Batey,R.T. (2009) A switch in time: detailing the life of a riboswitch. *Biochim. Biophys. Acta*, **1789**, 584–591.
5. Polaski,J.T., Holmstrom,E.D., Nesbitt,D.J. and Batey,R.T. (2016) Mechanistic insights into cofactor-dependent coupling of RNA folding and mRNA Transcription/Translation by a cobalamin riboswitch. *Cell Rep*, **15**, 1100–1110.
6. Caron,M.P., Bastet,L., Lussier,A., Simoneau-Roy,M., Masse,E. and Lafontaine,D.A. (2012) Dual-acting riboswitch control of translation initiation and mRNA decay. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, E3444–E3453.
7. Hollands,K., Proshkin,S., Sklyarova,S., Epshtein,V., Mironov,A., Nudler,E. and Groisman,E.A. (2012) Riboswitch control of Rho-dependent transcription termination. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 5376–5381.
8. Bastet,L., Chauvier,A., Singh,N., Lussier,A., Lamontagne,A.M., Prevost,K., Masse,E., Wade,J.T. and Lafontaine,D.A. (2017) Translational control and Rho-dependent transcription termination are intimately linked in riboswitch regulation. *Nucleic Acids Res.*, **45**, 7474–7486.
9. Schroeder,K.T., McPhee,S.A., Ouellet,J. and Lilley,D.M. (2010) A structural database for k-turn motifs in RNA. *RNA*, **16**, 1463–1468.
10. Chan,C.W., Chetnani,B. and Mondragon,A. (2013) Structure and function of the T-loop structural motif in noncoding RNAs. *Wiley Interdiscip. Rev. RNA*, **4**, 507–522.
11. Brunel,C., Marquet,R., Romby,P. and Ehresmann,C. (2002) RNA loop-loop interactions as dynamic functional motifs. *Biochimie*, **84**, 925–944.
12. Leontis,N.B. and Westhof,E. (1998) A common motif organizes the structure of multi-helix loops in 16 S and 23 S ribosomal RNAs. *J. Mol. Biol.*, **283**, 571–583.
13. Barrick,J.E. and Breaker,R.R. (2007) The distributions, mechanisms, and structures of metabolite-binding riboswitches. *Genome Biol.*, **8**, R239.
14. McCown,P.J., Corbino,K.A., Stav,S., Sherlock,M.E. and Breaker,R.R. (2017) Riboswitch diversity and distribution. *RNA*, **23**, 995–1011.
15. Butcher,S.E. and Pyle,A.M. (2011) The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. *Acc. Chem. Res.*, **44**, 1302–1311.
16. Ashraf,S., Huang,L. and Lilley,D.M.J. (2017) Sequence determinants of the folding properties of box C/D kink-turns in RNA. *RNA*, **23**, 1927–1935.
17. McPhee,S.A., Huang,L. and Lilley,D.M. (2014) A critical base pair in k-turns that confers folding characteristics and correlates with biological function. *Nat. Commun.*, **5**, 5127.
18. Schroeder,K.T., Daldrop,P. and Lilley,D.M. (2011) RNA tertiary interactions in a riboswitch stabilize the structure of a kink turn. *Structure*, **19**, 1233–1240.
19. Johnson,J.E. Jr., Reyes,F.E., Polaski,J.T. and Batey,R.T. (2012) B12 cofactors directly stabilize an mRNA regulatory switch. *Nature*, **492**, 133–137.
20. Polaski,J.T., Webster,S.M., Johnson,J.E. Jr. and Batey,R.T. (2017) Cobalamin riboswitches exhibit a broad range of ability to discriminate between methylcobalamin and adenosylcobalamin. *J. Biol. Chem.*, **292**, 11650–11658.
21. Lee,J.C., Cannone,J.J. and Gutell,R.R. (2003) The lonepair triloop: a new motif in RNA structure. *J. Mol. Biol.*, **325**, 65–83.
22. Lisi,V. and Major,F. (2007) A comparative analysis of the triloops in all high-resolution RNA structures reveals sequence structure relationships. *RNA*, **13**, 1537–1545.
23. Shaner,N.C., Lambert,G.G., Chammas,A., Ni,Y., Cranfill,P.J., Baird,M.A., Sell,B.R., Allen,J.R., Day,R.N., Israelsson,M. *et al.* (2013) A bright monomeric green fluorescent protein derived from Branchiostoma lanceolatum. *Nat. Methods*, **10**, 407–409.
24. Ceres,P., Garst,A.D., Marcano-Velazquez,J.G. and Batey,R.T. (2013) Modularity of select riboswitch expression platforms enables facile engineering of novel genetic regulatory devices. *ACS Synth. Biol.*, **2**, 463–472.
25. Geissmann,Q. (2013) OpenCFU, a new free and open-source software to count cell colonies and other circular objects. *PLoS One*, **8**, e54072.
26. Kearse,M., Moir,R., Wilson,A., Stones-Havas,S., Cheung,M., Sturrock,S., Buxton,S., Cooper,A., Markowitz,S., Duran,C. *et al.* (2012) Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.
27. Tamura,K. and Nei,M. (1993) Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. *Mol. Biol. Evol.*, **10**, 512–526.
28. Davis,J.H., Rubin,A.J. and Sauer,R.T. (2011) Design, construction and characterization of a set of insulated bacterial promoters. *Nucleic Acids Res.*, **39**, 1131–1141.
29. Lundrigan,M.D. and Kadner,R.J. (1989) Altered cobalamin metabolism in Escherichia coli btuR mutants affects btuB gene regulation. *J. Bacteriol.*, **171**, 154–161.
30. Baba,T., Ara,T., Hasegawa,M., Takai,Y., Okumura,Y., Baba,M., Datsenko,K.A., Tomita,M., Wanner,B.L. and Mori,H. (2006) Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.*, **2**, 2006.
31. Reetz,M.T., Kahakeaw,D. and Lohmer,R. (2008) Addressing the numbers problem in directed evolution. *ChemBioChem*, **9**, 1797–1804.
32. Edgar,R.C. (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Res.*, **32**, 1792–1797.
33. Randall,R.N., Radford,C.E., Roof,K.A., Natarajan,D.K. and Gaucher,E.A. (2016) An experimental phylogeny to benchmark ancestral sequence reconstruction. *Nat. Commun.*, **7**, 12847.
34. Hietpas,R.T., Jensen,J.D. and Bolon,D.N. (2011) Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci. U.S.A.*, **108**, 7896–7901.
35. Pitt,J.N. and Ferre-D'Amare,A.R. (2010) Rapid construction of empirical RNA fitness landscapes. *Science*, **330**, 376–379.
36. Fox,K.A., Ramesh,A., Stearns,J.E., Bourgogne,A., Reyes-Jara,A., Winkler,W.C. and Garsin,D.A. (2009) Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in Enterococcus faecalis. *Proc. Natl. Acad. Sci. U.S.A.*, **106**, 4435–4440.
37. Boniecki,M.J., Lach,G., Dawson,W.K., Tomala,K., Lukasz,P., Soltysinski,T., Rother,K.M. and Bujnicki,J.M. (2016) SimRNA: a coarse-grained method for RNA folding simulations and 3D structure prediction. *Nucleic Acids Res.*, **44**, e63.
38. Miao,Z., Adamiak,R.W., Antczak,M., Batey,R.T., Becka,A.J., Biesiada,M., Boniecki,M.J., Bujnicki,J.M., Chen,S.J., Cheng,C.Y.

et al. (2017) RNA-Puzzles Round III: 3D RNA structure prediction of five riboswitches and one ribozyme. *RNA*, **23**, 655–672.
39. Miao,Z. and Westhof,E. (2017) RNA Structure: Advances and assessment of 3D structure prediction. *Annu. Rev. Biophys.*, **46**, 483–503.
40. Huang,L. and Lilley,D.M.J. (2016) The kink turn, a key architectural element in RNA structure. *J. Mol. Biol.*, **428**, 790–801.
41. Jucker,F.M., Heus,H.A., Yip,P.F., Moors,E.H. and Pardi,A. (1996) A network of heterogeneous hydrogen bonds in GNRA tetraloops. *J. Mol. Biol.*, **264**, 968–980.
42. Burmann,B.M., Schweimer,K., Luo,X., Wahl,M.C., Stitt,B.L., Gottesman,M.E. and Rosch,P. (2010) A NusE:NusG complex links transcription and translation. *Science*, **328**, 501–504.
43. Proshkin,S., Rahmouni,A.R., Mironov,A. and Nudler,E. (2010) Cooperation between translating ribosomes and RNA polymerase in transcription elongation. *Science*, **328**, 504–508.
44. Garst,A.D., Porter,E.B. and Batey,R.T. (2012) Insights into the regulatory landscape of the lysine riboswitch. *J. Mol. Biol.*, **423**, 17–33.
45. Wickiser,J.K., Winkler,W.C., Breaker,R.R. and Crothers,D.M. (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell*, **18**, 49–60.
46. Woodson,S.A. (2014) Taming free energy landscapes with RNA chaperones. *RNA Biol.*, **7**, 677–686.
47. Stone,M.D., Mihalusova,M., O'Connor,C.M., Prathapam,R., Collins,K. and Zhuang,X. (2007) Stepwise protein-mediated RNA folding directs assembly of telomerase ribonucleoprotein. *Nature*, **446**, 458–461.
48. Li,C., Qian,W., Maclean,C.J. and Zhang,J. (2016) The fitness landscape of a tRNA gene. *Science*, **352**, 837–840.
49. Solomatin,S.V., Greenfeld,M., Chu,S. and Herschlag,D. (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature*, **463**, 681–684.
50. Weinberg,Z., Wang,J.X., Bogue,J., Yang,J., Corbino,K., Moy,R.H. and Breaker,R.R. (2010) Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. *Genome Biol.*, **11**, R31.
51. Leontis,N.B. and Westhof,E. (2001) Geometric nomenclature and classification of RNA base pairs. *RNA*, **7**, 499–512.
52. Schneider,T.D. and Stephens,R.M. (1990) Sequence logos: a new way to display consensus sequences. *Nucleic Acids Res.*, **18**, 6097–6100.
53. Crooks,G.E., Hon,G., Chandonia,J.M. and Brenner,S.E. (2004) WebLogo: a sequence logo generator. *Genome Res.*, **14**, 1188–1190.