B$_{12}$ cofactors directly stabilize an mRNA regulatory switch

James E. Johnson Jr.[*], **Francis E. Reyes**[*,†,1], **Jacob T. Polaski**, and **Robert T. Batey**[†]
Department of Chemistry and Biochemistry, University of Colorado at Boulder, UCB 215, Boulder, Colorado 80309-0215, USA

Abstract

Structures of riboswitch receptor domains bound to their effector have revealed how mRNAs recognize diverse small molecules, but mechanistic details into its linkage with regulation of gene expression remain elusive[1,2]. To address this, we solved crystal structures of two different classes of cobalamin (vitamin B$_{12}$) binding riboswitches that include the structural switch of the downstream regulatory domain. These classes share a common cobalamin-binding core, but use distinct peripheral extensions to recognize different B$_{12}$ derivatives. In each case, recognition is accomplished through shape complementarity between the RNA and cobalamin with relatively few hydrogen bonding interactions that typically govern RNA-small molecule recognition. We show that a composite cobalamin/RNA scaffold stabilizes an unusual long-range intramolecular kissing-loop interaction that controls mRNA expression. This is the first riboswitch crystal structure detailing how the receptor and regulatory domains communicate in a ligand-dependent fashion to regulate mRNA expression.

An mRNA leader identified as controlling expression of a cobalamin transport gene (*btuB*) in *E. coli*[3,4] was the first validated riboswitch shown to directly interact with cellular metabolites in the absence of proteins[5,6]. Since, cobalamin riboswitches have been found to widely regulate B$_{12}$ biosynthesis in bacteria[7,8] and are one of the most broadly distributed riboswitches in biology[9,10]. The cobalamin riboswitch family comprises two classes[5,8] distinguished by peripheral extensions surrounding a common core. The secondary structure of both classes contains a central four-way junction (P3–P6) forming the core receptor domain responsible for cobalamin binding (Fig. 1b; Supplementary Fig. 1). The other shared element is a kissing-loop (KL) interaction between L5 of the receptor and L13 of the regulatory domain that instructs the expression machinery. For cobalamin riboswitches that regulate translation, L13 typically contains the ribosome binding site (RBS). Within the first class, the KL is generally linked to a downstream secondary structural switch[8] akin to that

---

[†]Corresponding authors. Tel: (303) 735 2159, FAX: (303) 492 5894, robert.batey@colorado.edu, francis.reyes@colorado.edu.
[1]Current address: Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California, 94720, USA.
[*]These authors contributed equally to this work.

employed by most riboswitches[9]. The classes are primarily differentiated by the presence of a large peripheral extension of P6 in one class (representing ~95% of the sequences of the cobalamin family[10]) that is absent or severely truncated in the other class (Fig. 1b)[5,8]. Another peripheral extension between P1 and P3 further defines the classes. A unique feature of numerous cobalamin riboswitches, particularly those without the P6-extension, is that regulation appears to be achieved through tertiary structure formation (the kissing-loop interaction).

All cobalamin riboswitches are proposed to bind adenosylcobalamin (AdoCbl) despite experimental validation of only a few member sequences[5,6,11]. Unexpectedly, we found some riboswitches recognize methylcobalamin (MeCbl) and aquocobalamin (AqCbl) with significantly higher affinity than AdoCbl (Supplementary Fig. 2, Supplementary Table 1). Chemical probing[12] shows B_{12}-dependent reactivity changes in a riboswitch lacking the P6-extension only in the presence of derivatives with small β-axial moieties (Fig. 1a), unlike the *E. coli btuB* riboswitch that selectively binds adenosylcobalamin (Fig. 1c; Supplementary Fig. 3)[5,6]. Significant reactivity changes are localized to three regions: the central junction (J6/3), L5 and L13, the latter two suggestive of KL formation. Cobalamin selectivity and the role of KL formation in gene regulation was validated by cell-based assays. Control of reporter gene expression by a riboswitch lacking the P6-extension was tested in a Δ*btuR* (cobalamin adenosyl transferase) strain of *E. coli* incapable of converting AqCbl to AdoCbl[13]. GFPuv expression is repressed ~8.5-fold by AqCbl versus less than twofold with AdoCbl (Fig. 1d). A mutation of L5 in the mRNA (L5-GAAA) that retains cobalamin binding but eliminates KL formation abolishes repression, directly establishing the essential role of tertiary structure formation in regulating gene expression. Notably, similar sequences group on a branch of a cobalamin riboswitch phylogenetic tree highly populated by marine cyanobacterial[8] and environmental (*env*) metagenomes of samples from the ocean surface[14]. The prevalence of these riboswitches in these bacteria suggests adaptation of the RNA to environments where the free cobalamin pool is predominantly AqCbl due to rapid photolysis of AdoCbl[15]. Therefore, we refer to riboswitches containing the P6-extension as "AdoCbl" and those lacking this extension as "AqCbl", reflecting their likely biological effectors.

To determine the mechanistic basis for cobalamin-dependent regulation we solved structures of the *env8*AqCbl riboswitch in complex with AqCbl (Fig. 2a) and the *Thermoanaerobacter tengcongensis* (*Tte*) AdoCbl riboswitch bound to AdoCbl (Fig. 2b) (see Full Methods). AdoCbl riboswitches are the largest of the known riboswitches—at over 200 nucleotides they are the size of the *Azoarcus* group I self-splicing intron[16]. The *env8*AqCbl riboswitch is the first structure containing both the receptor and regulatory domain. Insertion of the sequence spanning the crystal structure upstream of a GFPuv reporter confers cobalamin-dependent regulatory activity, albeit at lower efficiency than the wild-type riboswitch (xtal, Fig. 1d). Thus, the *env8*AqCbl structure corresponds to a completely functional riboswitch encompassing all of the sequence necessary and sufficient to impart biological activity.

The global architecture of both RNAs is defined by organization of their common secondary structure into two coaxial stacks, P1/P3/P6 and P4/P5/P13 (blue and green, Fig. 2), consistent with comparative analysis of known RNA structures[17] and mutual information sequence analysis[9]. These stacks are joined by a T-loop/T-loop motif (L4-L6), a common

module of RNA tertiary architecture[18,19]. In *Tte*AdoCbl, the T-loop (L4) interacts with an internal loop between P6 and P7 that partially mimics the structure of the T-loop (Supplementary Fig. 4), rather than another T-loop as observed in the AqCbl and FMN riboswitches[20]. Class-specific peripheral extensions flank the core (magenta and cyan, Fig. 2) that contact J6/3, a key element for cobalamin recognition.

Interactions between the RNA and cobalamin are mediated primarily through van der Waals shape complementarity with few direct hydrogen bonds. In both structures, cobalamin is sandwiched between the minor grooves of P3/P6 and the helix created by base pairing of L5 and L13, forming the regulatory KL (Fig. 2). The J3/4 and J6/3 strands are central to the receptor side of the binding pocket; in *env8*AqCbl, stacking of four purines from these strands creates a relatively flat surface (G19, A20, A67 and A68; Fig. 3a). Cobalamin's β-axial face projects directly toward this surface so the plane of the corrin ring is almost perpendicular to the bases of the purine stack. Despite numerous propionamide and acetamide groups surrounding the corrin ring, only one acetamide contacts the minor groove edge of G19 in *env8*AqCbl and its equivalent in *Tte*AdoCbl (G49) (Fig. 3b, c). A SELEX-generated aptamer uses a similar strategy to bind cyanocobalamin, although the molecular details of recognition are quite different (Supplementary Fig. 5)[21].

Selectivity between cobalamin derivatives is achieved through conformational differences in J6/3 mediated by the peripheral extensions. In *env8*AqCbl, proximity of A20 and A68 to the corrin ring sterically occludes the 5'-deoxyadenosyl moiety of AdoCbl, establishing its selectivity for cobalamins with small β-axial moieties (Fig. 3b). The conformation of J6/3 that blocks AdoCbl binding is enforced through base pairing with the J1/3 peripheral extension (G10•U69 and C11-G70). In *Tte*AdoCbl, placement of J6/3 further from J3/4 allows the Hoogsteen face of A162, the equivalent of A68 in *env8*AqCbl, to base pair with the 5'-deoxyadenosyl moiety (Fig. 3c). Positioning of J6/3 is reinforced by two highly conserved adenosines (A130 and A131) in the internal loop between P10 and P11 of the P6-extension (Fig. 3c, Supplementary Fig. 6), in support of the finding that this extension is required for AdoCbl specificity (Supplementary Table 1). These adenosines are strongly protected from chemical modification only in the presence of AdoCbl (J11/10, Supplementary Fig. 3), suggesting that the peripheral domain docks with the core after initial cobalamin binding. Since other cobalamins can bind the AdoCbl riboswitch with ~80-fold lower affinity (Supplementary Table 1), we hypothesize that the P6-extension may allow these riboswitches to differentially regulate gene expression based upon whether the intracellular cobalamin pool is dominated by AdoCbl or its photolysed product AqCbl. Intriguingly, a recently discovered repressor protein in *Myxococcus xanthus* uses B$_{12}$ to photoregulate carotenoid biosynthesis[22]. Use of peripheral extensions by a riboswitch to modulate its properties has not been previously observed, but is common with other large RNAs[23,24].

The receptor-cobalamin complex presents a composite surface composed of RNA and ligand that binds the regulatory domain. Superimposition of the full *env8*AqCbl riboswitch with a structure of the receptor domain alone indicates that the receptor-ligand complex provides a relatively rigid surface for docking of P13 (Supplementary Fig. 7). RNA-RNA interactions between the two domains are through base pairing of nucleotides in L5 and L13 to form the

KL. In both structures, the α-axial face of B_{12} interacts with the KL through structure-specific rather than sequence-specific contacts. Dimethylbenzimidazole and the aminopropyl linker interact with the ribose-phosphate backbone of L5 and L13 primarily through van der Waals contacts; the only direct hydrogen bond is between the 2'-hydroxyl group of the cobalamin ribosyl moiety and the 2'-hydroxyl group of C91/C194 (Fig. 4a). Further contacts are made to the RNA by propionamide and acetamide groups with the only direct base contact being to the sugar edge of U44 (*env8*AqCbl) or its equivalent, G74 (*Tte*AdoCbl), in L5. The KL-cobalamin interface is nearly identical between the two structures despite low sequence similarity in L5 and L13, indicating structure-specific recognition of the KL by B_{12}.

Kissing-loops are inherently stable structures widely used to promote tertiary architecture formation and RNA-RNA interactions[25,26]. A unique feature of the KL in both structures is the presence of non-Watson-Crick pairs and bulged nucleotides in the helix formed by pairing of L5 and L13 ("MM" and "B", Fig. 4a). Mismatched pairs are strongly destabilizing to the KL[27], suggesting a mechanism for creating a cobalamin-dependent switch. To show bound cobalamin promotes KL formation, its structure was probed as a function of Mg^{2+}, which promotes long-range tertiary interactions in RNA[28]. Chemical probing of the unbound AqCbl riboswitch shows that at least 15 mM Mg^{2+} is required before reactivity protections consistent with KL formation are observed, whereas bound ligand allows its formation under physiological Mg^{2+} concentrations (0.5–1 mM) (Fig. 4b, Supplementary Fig. 8). This same trend is observed for the *E. coli btuB* AdoCbl riboswitch (Supplementary Fig. 9). Stabilizing the KL with mutations in L13 that enable perfect Watson-Crick pairing with L5 (G95U, ΔA96, ΔG97) results in cobalamin-independent KL formation, equivalent to a constitutively repressed mRNA due to sequestration of the RBS (Fig. 4c). Together, these data reveal that cobalamin riboswitches employ ligand-dependent formation of a tertiary RNA module as the basis for gene regulation. This is a structurally distinct but functionally equivalent mechanism to the majority of other RNA switches that use mutually exclusive secondary structures to achieve regulatory activity.

Full Methods

RNA Preparation

RNA constructs (full sequences are shown in Supplemental Fig. 1) were prepared using DNA templates generated from PCR amplification using established protocols[30]. DNA templates for the *E. coli btuB* cobalamin riboswitch was amplified from *E. coli* genomic DNA, while the *env4*, *env8* and *T. tengcongensis* cobalamin riboswitches were amplified using a series of overlapping oligonucleotides. PCR products were used as template for transcription reactions using T7 RNA polymerase[31] and the RNA purified using denaturing PAGE (8% or 12%, 29:1 acrylamide:bisacrylamide). The RNA was buffer exchanged and concentrated into 0.5x T.E. buffer and frozen at -80 °C until use.

Isothermal Titration Calorimetry

RNA was dialyzed overnight into a buffer containing 5 mM NaMES, pH 6.0, 100 mM KCl, and 5 mM $MgCl_2$. RNA was diluted up to a final concentration of 10 μM and titrated with

either AdoCbl, AqCbl or MeCbl dissolved in the dialysis buffer at concentrations 10-fold in excess of the RNA. Titrations were all performed at 25 °C using a MicroCal iTC$_{200}$ microcalorimeter. Data analysis and fitting was performed with the Origin software suite as previously described[32].

Chemical probing with *N*-methylisatoic acid

Chemical probing of RNAs were performed using slight modifications to established protocols[12]. Purified RNA constructs were refolded by incubation at 70 °C for three minutes, room temperature for five minutes, and on ice for a minimum five minutes. For ligand comparison experiments, 10 μL solutions were prepared to a final concentration of 0.1 μM RNA, 100 mM K·HEPES, pH 8.0, 100 mM NaCl, 6 mM MgCl$_2$, ligand, and 6.5 mM *N*-methylisatoic acid (NMIA). Final ligand concentrations for the *E. coli btuB* experiments were 0.5 mM AdoCbl, 2.5 mM AqCbl, CNCbl, and MeCbl. For the *env4*AqCbl experiments, final ligand concentrations were 100 μM for all compounds. For magnesium titrations, the conditions were the same as the above mentioned conditions with the final concentrations of ligands as 500 μM AdoCbl or 100 μM MeCbl. Samples were reverse transcribed as previously described[12]. Products were separated using 12% denaturing polyacrylamide gel and visualized using a Typhoon PhosphoroImager (Molecular Dynamics).

In vivo reporter assay

For all *in vivo* assays, *E. coli* Δ*btuR* (Keio collection[33] JW1262) cells were grown in a rich, chemically defined medium that was supplemented with (100 μg/mL) ampicillin and a varying amount of a cobalamin. For titration experiments, 5 μL of overnight culture was added to 5 mL of media and incubated for 6 hours at 37 °C. Fluorescence and OD$_{600}$ measurements were performed on 300 μL of cells from each replicate in clear-bottom 96-well plates. GFPuv fluorescence was read at an excitation wavelength of 395 nm and a 510 nm emission wavelength using an Xfluor SafireII fluorimeter (Tecan). All data shown represent average fluorescence values of three biological replicates that were normalized to the OD$_{600}$ in each well. All fluorescence measurements were background corrected by taking identical fluorescence and OD$_{600}$ measurements for *E. coli* Δ*btuR* cells transformed with pBR322 vector that did not harbor the reporter gene. Background fluorescence was subtracted from total fluorescence and fold-repression was calculated by dividing the average normalized/background corrected fluorescence values for the unrepressed construct (−Cbl) by the average normalized/background corrected fluorescence value for each repressed construct (+Cbl). Titration data was fit to a two state equation to determine EC$_{50}$ and error bars represent the standard deviation for each triplicate fluorescence measurement.

Structure solution of the *env8*AqCbl(ΔJ1/13,P13)/AqCbl complex

A representative sequence from the original sequence alignment (Genbank Accession #AACY021350931.1/557-442) was selected. The P13 stem was removed and the anti-RBS sequence in P5 was replaced with a stable GAAA tetraloop. The RNA was synthesized and refolded as described above. Diffraction quality crystals grew within 3 days at 30 °C in the presence of 100 mM magnesium acetate, 10% 2-methyl-1,3 propanediol (MPD), and 5 mM

iridium hexammine. Crystals were cryoprotected in Ficoll oil and flash frozen in liquid nitrogen. Complete datasets at the iridium and cobalt anomalous edges were collected and processed via autoPROC.[34] The structure was solved using PHENIX[35]. The density modified map showed excellent density for the entire RNA (Supplementary Fig. 10a, b). The model was built manually in COOT[36] and refined with BUSTER[37].

Structure solution of the *env8*AqCbl/AqCbl complex

The original *env8*AqCbl riboswitch (Genbank Accession #: AACY021350931.1/557-442) was modified with the following mutations: G12A, A14G, A31U, G42C, C62G and the replacement of the linker region with that of a sequence from the original alignment (Genbank Accession #: AACY023653040/384-265). Each RNA synthesized was refolded as described above. Crystals grew within 3 months at 20 °C under 10% v/v MPD, 40 mM sodium cacodylate pH 7.0, 12 mM spermine tetrahydrochloride, 80 mM KCl, and 20 mM BaCl$_2$ via hanging drop vapour diffusion. A complete dataset was collected on a crystal flash frozen in liquid nitrogen at the cobalt anomalous edge and reduced with autoPROC[34]. The structure was solved via molecular replacement using *env8*AqCbl(ΔJ1/13,P13)/AqCbl. The solution was confirmed by anomalous difference fourier maps showing the location of aquocobalamin. BUSTER[37] refinement produced clear density for P5 and P13 (Supplementary Fig. 10c, d). Model building was assisted with RCRANE[38] within COOT.[36] The final model was prepared using restrained refinement within BUSTER[37]. Data collection, phasing and refinement statistics for the *env8*AqCbl/AqCbl and *env8*AqCbl(ΔJ1/13,P13)/AqCbl can be found in Supplementary Table 2.

Structure solution of the *Tte*AdoCbl/AdoCbl complex

The sequence of the original *Tte*AdoCbl riboswitch (Genbank Accession #: AE008691.1/395133-395373) was modified to substitute the human U1A binding protein RNA motif[39] and a stable GAAA closing tetraloop for the nonconserved P2 and P10 helices, respectively. The RNA was synthesized and refolded as described above, to which 2 μl of a 4 mM solution of AdoCbl was added. Crystals grew over the course of 3 days at 30 °C via the hanging drop method in 10% isopropanol, 300 mM MgCl$_2$, and 100 mM Na-HEPES pH 7.5 in the absence of U1A and cryoprotected in 30% glycerol. Heavy atom derivatives were prepared by including 2–10 mM compound in the mother liquor during crystallization.

All datasets were collected at the cobalt anomalous edge for native datasets and the anomalous edges for TaBr and Ir derivatives. Molecular replacement in PHASER[40] using *env8*AqCbl RNA yielded a solution sufficient to locate heavy atoms in all datasets using anomalous difference data. Cross-crystal dispersive differences for cobalt were used to pair native datasets with heavy atom derivatives to create three sets of experimental phases (Supplementary Table 3). The phases were combined via DMMULTI resulting in traceable density maps. Iterative cycles of model building, refinement in BUSTER and PHENIX, multiple crystal averaging and heavy atom phasing, allowed most of the RNA to be built; the U1A loop at the apex of P2, the junction between P8-P10, and J1/13 remain unresolved in the electron density. Supporting electron density maps are shown in Supplementary Fig. 12. Data collection, phasing and refinement statistics can be found in Supplementary Table 3.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

References

1. Garst AD, Edwards AL, Batey RT. Riboswitches: structures and mechanisms. Cold Spring Harb Perspect Biol. 2011; 3

2. Breaker RR. Riboswitches and the RNA world. Cold Spring Harb Perspect Biol. 2012; 4

3. Lundrigan MD, Koster W, Kadner RJ. Transcribed sequences of the *Escherichia coli btuB* gene control its expression and regulation by vitamin B_{12} . Proc Natl Acad Sci U S A. 1991; 88:1479–1483. [PubMed: 1847525]

4. Franklund CV, Kadner RJ. Multiple transcribed elements control expression of the *Escherichia coli btuB* gene. J Bacteriol. 1997; 179:4039–4042. [PubMed: 9190822]

5. Nahvi A, Barrick JE, Breaker RR. Coenzyme B_{12} riboswitches are widespread genetic control elements in prokaryotes. Nucleic Acids Res. 2004; 32:143–150. [PubMed: 14704351]

6. Nahvi A, et al. Genetic control by a metabolite binding mRNA. Chemistry & Biology. 2002; 9:1043. [PubMed: 12323379]

7. Rodionov DA, Vitreschak AG, Mironov AA, Gelfand MS. Comparative genomics of the vitamin B_{12} metabolism and regulation in prokaryotes. J Biol Chem. 2003; 278:41148–41159. [PubMed: 12869542]

8. Vitreschak AG, Rodionov DA, Mironov AA, Gelfand MS. Regulation of the vitamin B_{12} metabolism and transport in bacteria by a conserved RNA structural element. RNA. 2003; 9:1084–1097. [PubMed: 12923257]

9. Barrick JE, Breaker RR. The distributions, mechanisms, and structures of metabolite-binding riboswitches. Genome Biol. 2007; 8:R239. [PubMed: 17997835]

10. Gardner PP, et al. Rfam: updates to the RNA families database. Nucleic Acids Res. 2009; 37:D136–D140. [PubMed: 18953034]

11. Fox K, et al. Multiple posttranscriptional regulatory mechanisms partner to control ethanolamine utilization in *Enterococcus faecalis*. Proc Natl Acad Sci USA. 2009

12. Wilkinson KA, Merino EJ, Weeks KM. Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc. 2006; 1:1610–1616. [PubMed: 17406453]

13. Lundrigan MD, Kadner RJ. Altered cobalamin metabolism in *Escherichia coli btuR* mutants affects *btuB* gene regulation. J Bacteriol. 1989; 171:154–161. [PubMed: 2644187]

14. Weinberg Z, et al. Comparative genomics reveals 104 candidate structured RNAs from bacteria, archaea, and their metagenomes. Genome Biol. 2010; 11:R31. [PubMed: 20230605]

15. Taylor RT, Smucker L, Hanna ML, Gill J. Aerobic photolysis of alkylcobalamins: quantum yields and light-action spectra. Arch Biochem Biophys. 1973; 156:521–533. [PubMed: 4718782]

16. Adams PL, Stahley MR, Kosek AB, Wang J, Strobel SA. Crystal structure of a self-splicing group I intron with both exons. Nature. 2004; 430:45–50. [PubMed: 15175762]

17. Jaeger L, Verzemnieks EJ, Geary C. The UA_handle: a versatile submotif in stable RNA architectures. Nucleic Acids Res. 2009; 37:215–230. [PubMed: 19036788]

18. Nagaswamy U, Fox GE. Frequent occurrence of the T-loop RNA folding motif in ribosomal RNAs. RNA. 2002; 8:1112–1119. [PubMed: 12358430]

19. Krasilnikov AS, Mondragon A. On the occurrence of the T-loop RNA folding motif in large RNA molecules. RNA. 2003; 9:640–643. [PubMed: 12756321]

20. Serganov A, Huang L, Patel DJ. Coenzyme recognition and gene regulation by a flavin mononucleotide riboswitch. Nature. 2009; 458:233–237. [PubMed: 19169240]

21. Sussman D, Nix JC, Wilson C. The structural basis for molecular recognition by the vitamin B_{12} RNA aptamer. Nat Struct Biol. 2000; 7:53–57. [PubMed: 10625428]

22. Ortiz-Guerrero JM, Polanco MC, Murillo FJ, Padmanabhan S, Elias-Arnanz M. Light-dependent gene regulation by a coenzyme B12-based photoreceptor. Proc Natl Acad Sci U S A. 2011; 108:7565–7570. [PubMed: 21502508]

23. Vicens Q, Cech TR. Atomic level architecture of group I introns revealed. Trends Biochem Sci. 2006; 31:41–51. [PubMed: 16356725]

24. Melnikov S, et al. One core, two shells: bacterial and eukaryotic ribosomes. Nat Struct Mol Biol. 2012; 19:560–567. [PubMed: 22664983]

25. Xin Y, Laing C, Leontis NB, Schlick T. Annotation of tertiary interactions in RNA structures reveals variations and correlations. RNA. 2008; 14:2465–2477. [PubMed: 18957492]

26. Butcher SE, Pyle AM. The molecular interactions that stabilize RNA tertiary structure: RNA motifs, patterns, and networks. Acc Chem Res. 2011; 44:1302–1311. [PubMed: 21899297]

27. Gregorian RS Jr, Crothers DM. Determinants of RNA hairpin loop-loop complex stability. J Mol Biol. 1995; 248:968–984. [PubMed: 7539081]

28. Pyle AM. Metal ions in the structure and function of RNA. J Biol Inorg Chem. 2002; 7:679–690. [PubMed: 12203005]

29. Reyes FE, Garst AD, Batey RT. Strategies in RNA crystallography. Methods Enzymol. 2009; 469:119–139. [PubMed: 20946787]

30. Edwards AL, Garst AD, Batey RT. Determining structures of RNA aptamers and riboswitches by X-ray crystallography. Methods Mol Biol. 2009; 535:135–163. [PubMed: 19377976]

31. Milligan JF, Uhlenbeck OC. Synthesis of small RNAs using T7 RNA polymerase. Methods Enzymol. 1989; 180:51–62. [PubMed: 2482430]

32. Gilbert SD, Batey RT. Monitoring RNA-ligand interactions using isothermal titration calorimetry. Methods Mol Biol. 2009; 540:97–114. [PubMed: 19381555]

33. Baba T, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006; 2 2006 0008.

34. Vonrhein C, et al. Data processing and analysis with the autoPROC toolbox. Acta Crystallographica Section D Biological Crystallography. 2011; 67:293–302.

35. Adams PD, et al. PHENIX: building new software for automated crystallographic structure determination. Acta Crystallographica Section D Biological Crystallography. 2002; 58:1948–1954.

36. Emsley P, Lohkamp B, Scott WG, Cowtan K. Features and development of Coot. Acta Crystallographica Section D Biological Crystallography. 2010; 66:486–501.

37. Blanc E, et al. Refinement of severely incomplete structures with maximum likelihood in BUSTER-TNT. Acta Crystallographica Section D Biological Crystallography. 2004; 60:2210–2221.

38. Keating KS, Pyle AM. Semiautomated model building for RNA crystallography using a directed rotameric approach. Proceedings of the National Academy of Sciences of the United States of America. 2010; 107:8177–8182. [PubMed: 20404211]

39. Ferré-D'Amaré AR. Use of the spliceosomal protein U1A to facilitate crystallization and structure determination of complex RNAs. Methods (San Diego, Calif). 2010; 52:159–167.

40. McCoy AJ, et al. Phaser crystallographic software. Journal of Applied Crystallography. 2007; 40:658–674. [PubMed: 19461840]
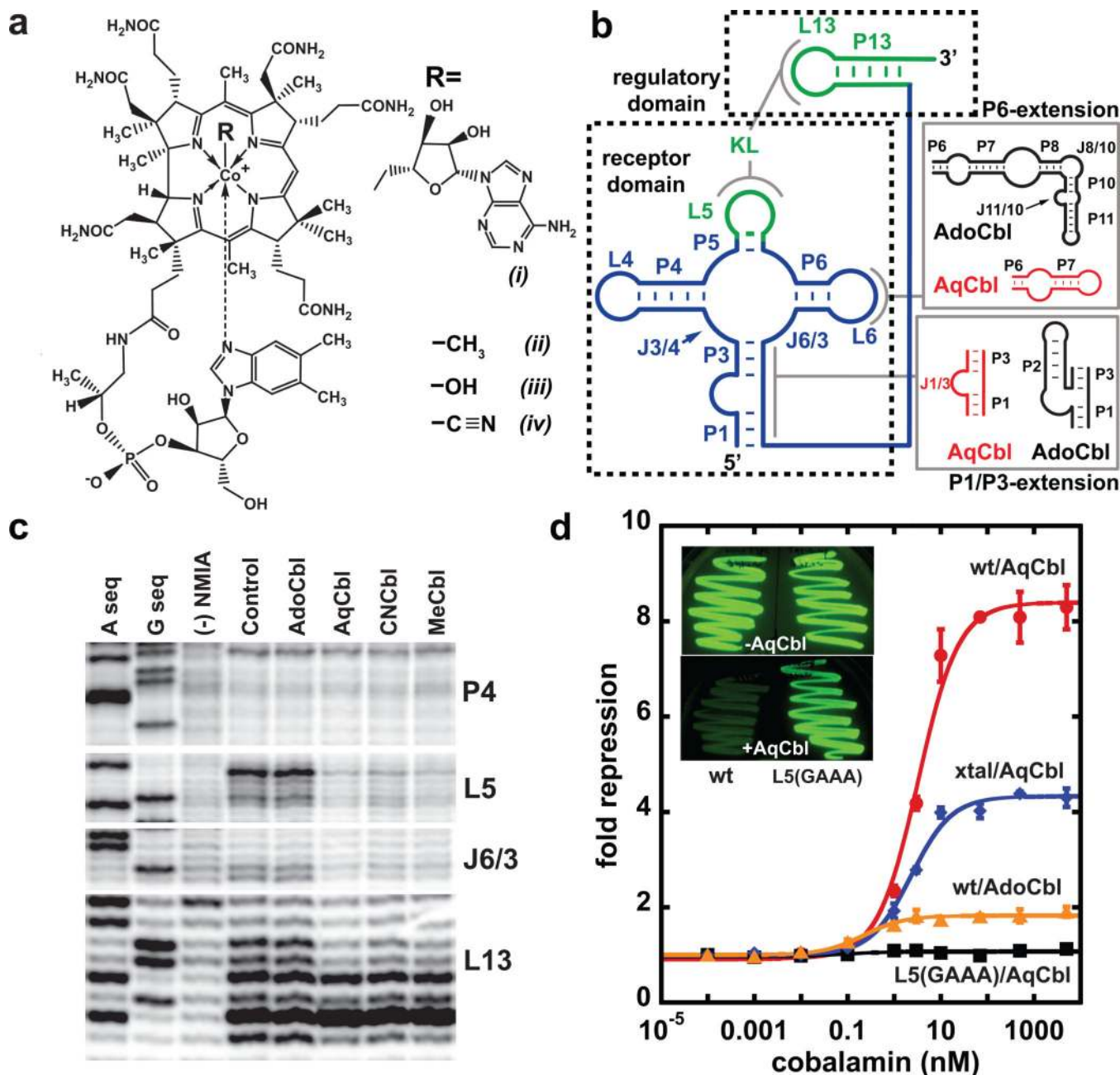
Figure 1. Structures of cobalamins and cobalamin riboswitches

(a) Cobalamins contain a corrin ring with a cobalt atom coordinated by an α-axial dimethylbenzimidazole (DMB) and by a variable β-axial group (R: adenosylcobalamin (AdoCbl, *i*), methylcobalamin (MeCbl, *ii*), aquocobalamin (AqCbl, *iii*), or cyanocobalamin (CNCbl, *iv*)). (b) Secondary structure of the cobalamin riboswitch family. The conserved core is shown in blue, the kissing-loop interaction in green, and peripheral extensions distinguishing the two classes are shown in black (AdoCbl) and red (AqCbl). (c) Chemical probing of the *env4*AqCbl riboswitch in the presence of cobalamins. A and G sequencing lanes are shown to the left followed by no probing reagent or ligand controls. (d) Repression of GFPuv expression by the *env8*AqCbl riboswitch in *E. coli*. Red and orange circles denote

wild-type *env8*AqCbl riboswitch in the presence of AqCbl and AdoCbl, respectively, blue diamonds a leader sequence that spans the crystallized riboswitch sequence and black squares an L5 mutant that cannot form the KL with L13. Error bars represent ± 1 s.d. The inset shows plates of the wild-type and L5 mutant grown in the absence or presence of AqCbl.
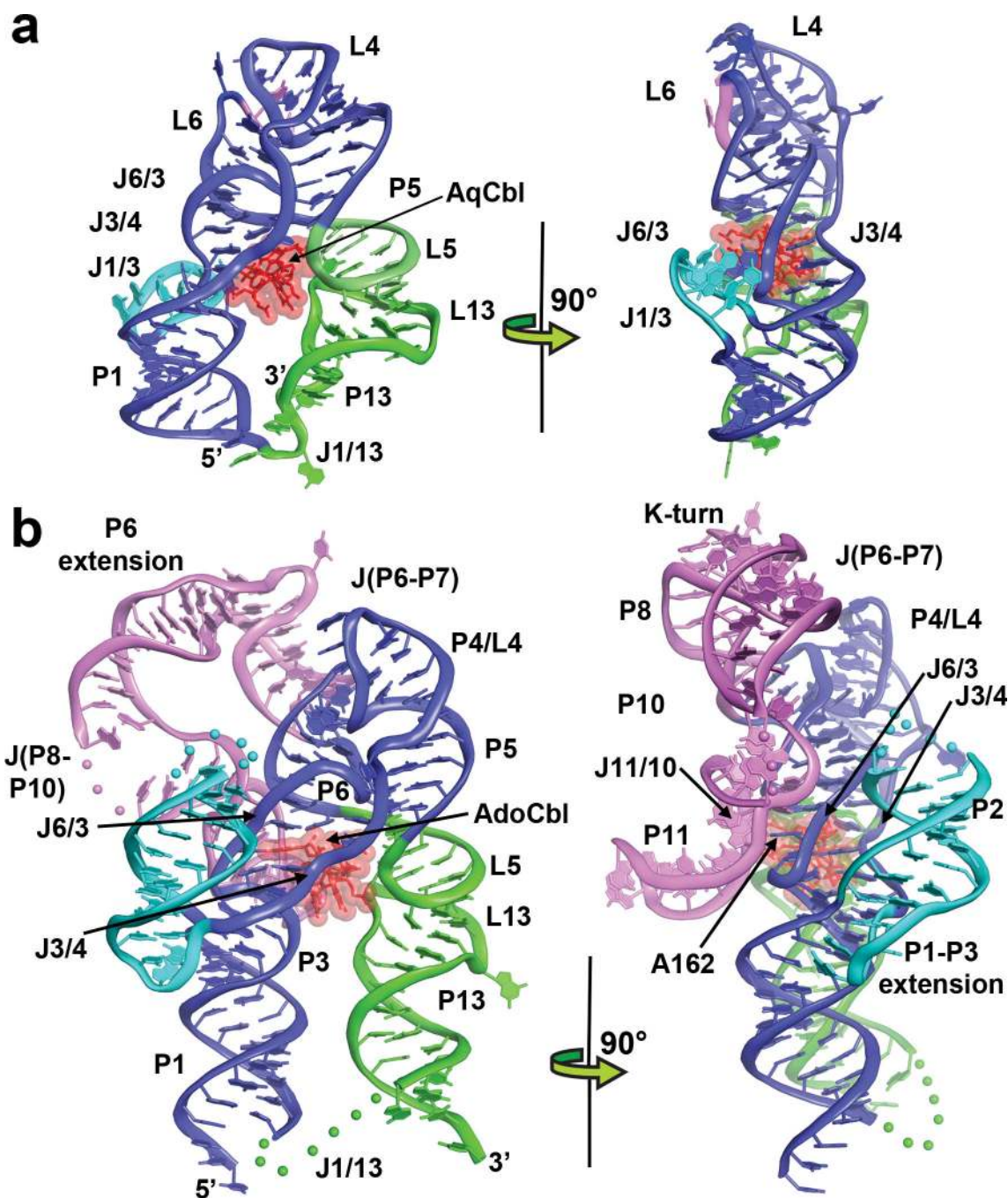
Figure 2. Crystal structures of two distinct cobalamin riboswitches

(a) Structure of the *env8*AqCbl riboswitch from a "front" and "side" perspective. Colored regions represent the core aptamer domain (blue), regulatory kissing-loop interaction (KL, green) and peripheral subdomains (magenta and cyan). AqCbl is represented by van der Waals spheres (red). A secondary structural representation of *env8*AqCbl RNA reflecting the tertiary organization and non-canonical base pairing is shown in Supplementary Fig. 11. (b) Cartoon representation of the structure of the *Tte*AdoCbl riboswitch complex. Coloring

of the RNA and ligand is consistent with panel (a). Disordered regions L2, J(P8-P10) and J1/13 are represented by spheres connecting the ends of the chain break.
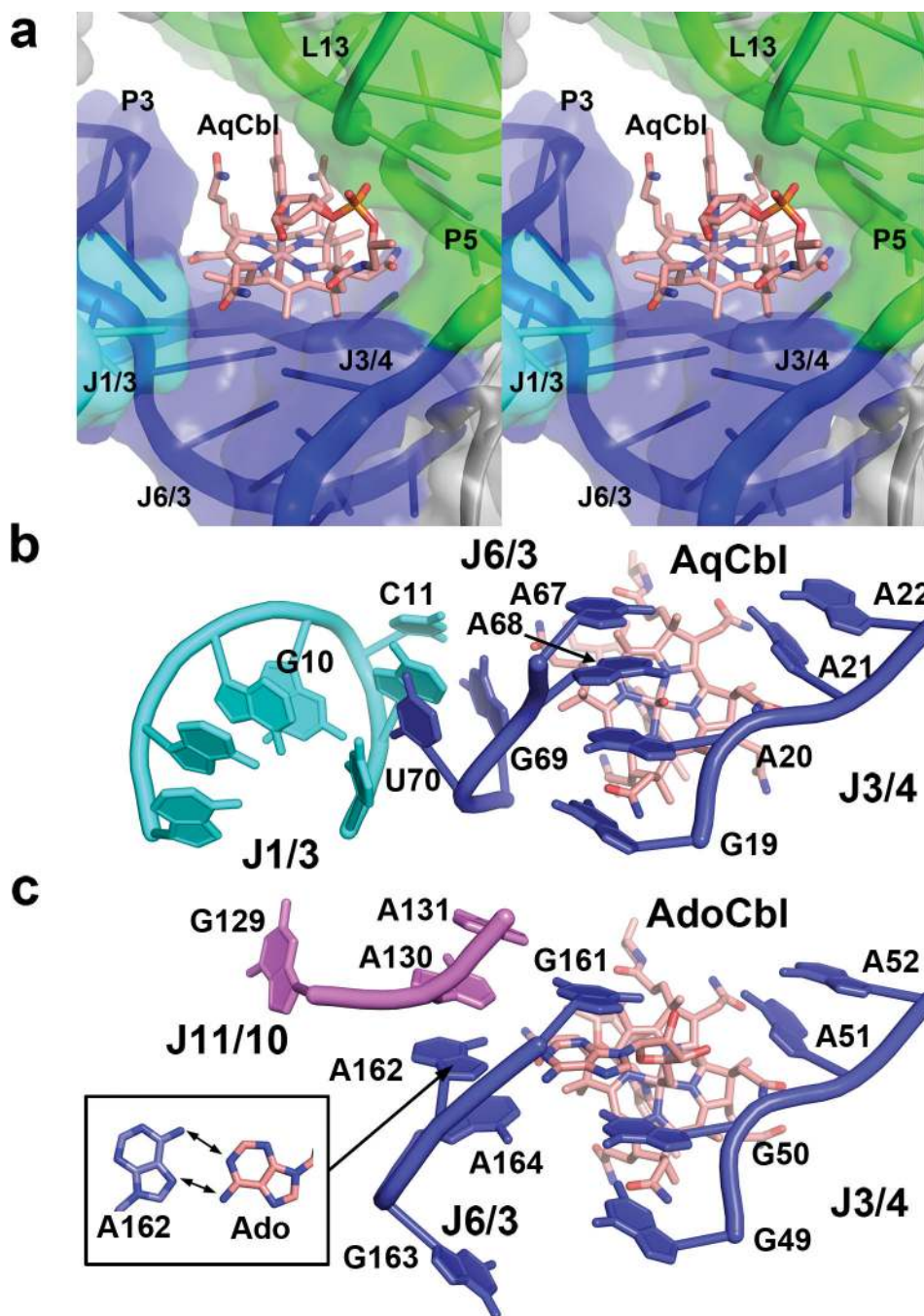
Figure 3. Cobalamin recognition by the receptor domain

(a) Stereoview of the surface representation of cobalamin binding to the receptor domain, emphasizing extensive shape complementarity between RNA and AqCbl. Coloring is consistent with Fig. 2. (b) Recognition of AqCbl by the *env8*AqCbl riboswitch. A stack of four purines in J3/4 and J6/3 packs against the β-axial face of the corrin ring with J6/3 directly buttressed by J1/3 (cyan). (c) Recognition of AdoCbl by the *Tte*AdoCbl RNA. While J3/4 is positioned identically to AqCbl, J6/3 adopts a different configuration that allows A162 to pair with the adenosyl base of AdoCbl (inset).
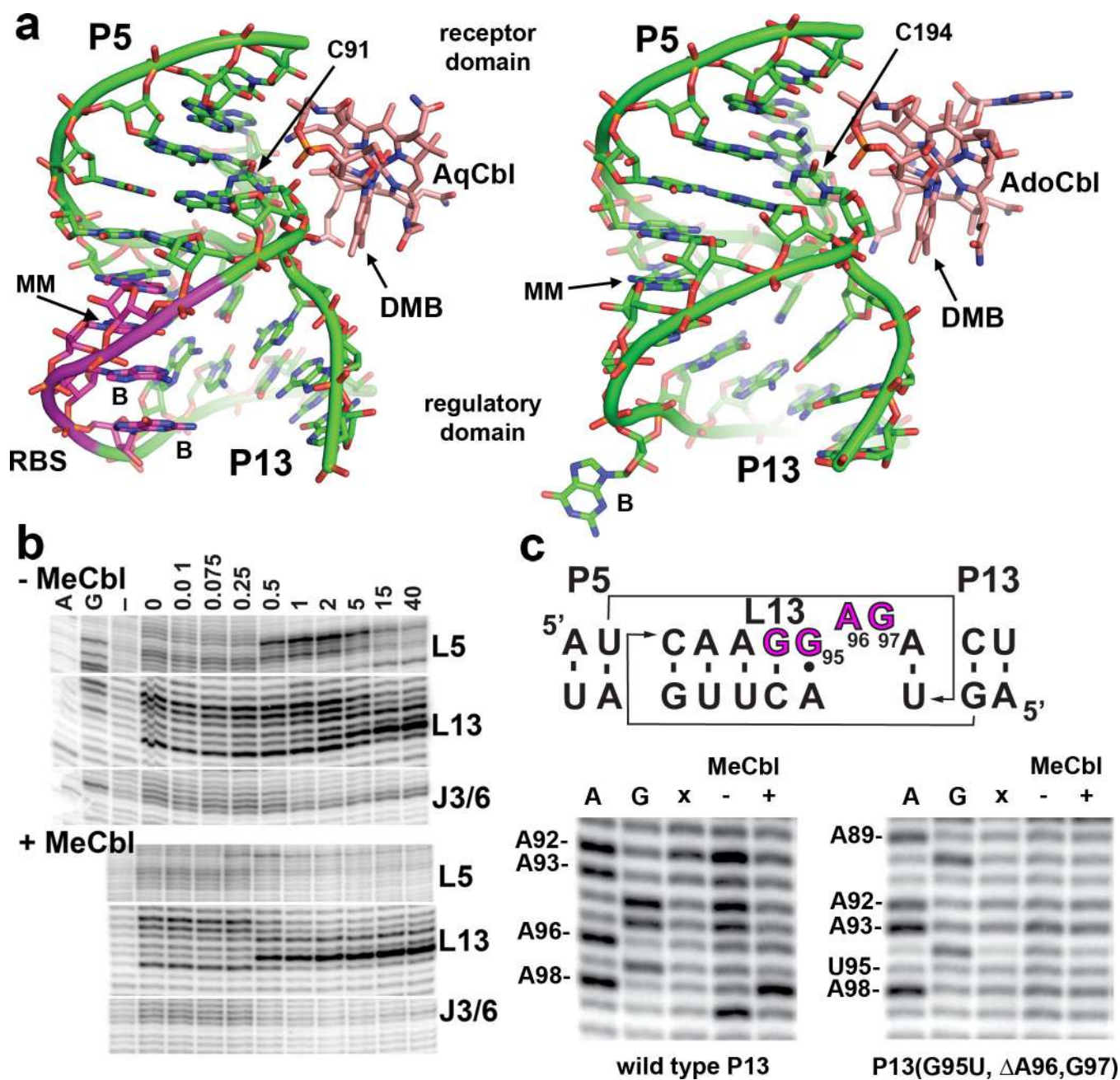
Figure 4. Formation of the kissing-loop interaction is the basis of cobalamin-dependent regulatory activity

(a) Detailed view of the cobalamin-KL interaction for the *env8*AqCbl (left) and *Tte*AdoCbl (right). Orientation of cobalamin is the same in the two perspectives, highlighting the interaction of the DMB moiety and one side of the corrin ring with the KL. The putative ribosome binding site (RBS) is highlighted in magenta, as well as base mismatches (MM) and bulged nucleotides (B) in the kissing-loop. (b) Chemical probing of the *env4*AqCbl riboswitch in the absence or presence of 100 μM MeCbl with increasing $MgCl_2$ concentrations. Sequencing and controls are the same as in Fig. 1. (c) Chemical probing of wild-type *env4*AqCbl (left) and a triple mutant (right) creating a perfectly complementary

P5/13 helix. Sequencing and controls are the same as in Fig. 1, followed by lanes corresponding to the absence (−) and presence (+) of MeCbl. KL sequence and pairing is shown above the gels for reference, highlighting the proposed RBS highlighted in magenta.