Supplemental Information

Mechanistic Insights into Cofactor-Dependent Coupling of RNA Folding and mRNA Transcription/Translation by a Cobalamin Riboswitch

Jacob T. Polaski, Erik D. Holmstrom, David J. Nesbitt, and Robert T. Batey
Figure S1. Cell density normalized GFPuv expression of riboswitches carrying stabilizing or destabilizing mutations in L5, related to Figure 2. (A) Sequence of the wild type KL and mutants with insertions or deletions in L5. Stabilizing mutations introduced into L5 of the wild type riboswitch are colored red and the L5(UGAAAG) mutation that ablates regulatory activity is colored green. (B) Cell density normalized fluorescence of cells expressing the RNAs depicted in A grown in the absence (gray bars) or presence (red bars) of 5 µM HyCbl. GFPuv expression levels from cells expressing the wild type riboswitch are shown as striped bars, and the inactive L5(UGAAAG) mutant is shown at the far right. Data are represented as mean ± SEM of triplicate measurements from at least three biological replicates.
Figure S2. RNAs with mutations in L5 bind hydroxocobalamin \textit{in vitro}, related to Figure 2. (A) Representative ITC thermograms of the wild type riboswitch and the L5(UGAAAG) mutant. (B) Table of binding affinities of the wild type RNA and the L5(UGAAAG) mutant for HyCbl. ITC traces of the wild type sequence were fit to a two-state binding model that revealed high and low affinity binding modes. Data are represented as mean ± SD of three independent titrations. (C) Representative ITC thermogram of a riboswitch carrying mutations in L5. (D) Table of binding affinities for HyCbl of the mutants depicted in Figures 2A and 2B. Error bars represent the mean ± SD of three independent titrations.
Figure S3. Regulatory activity is impacted by base pairing in the P13 helix, related to Figure 2. (A) Fold repression of GFPuv in cells expressing riboswitches with mutations in P13. Sequence schematics of P13 for each variant are shown above the graph, where mutations predicted to destabilize P13 are shown in red and compensatory mutations designed to restore helix formation are shown in green. Fold repression of cells expressing the wild type (wt) riboswitch is shown at the left as a striped bar, and the L5(UGAAAG) (ko) mutant that is unable to regulate gene expression is shown at the far right. Numbers denoting the location of each mutation correspond to the secondary structure of the env8HyCbl riboswitch shown in Figure 1A. Fold repression was calculated in the same manner described in Figure 2B. Values of fold repression for the wild type riboswitch and L5(UGAAAG) mutant are shown as mean ± SEM of triplicate measurements from at least three biological replicates. Fold repression for the P13 mutants was calculated using measurements from a single biological replicate. (B) Cell density normalized fluorescence of cells expressing the P13 mutants depicted in A grown in the absence (gray bars) or presence (red bars) of 5 µM HyCbl. Cell density normalized fluorescence from cells expressing the wild type (wt) riboswitch is shown at the left as striped bars and the L5(UGAAAG) (ko) variant is shown at the far right. For the wild type riboswitch and L5(UGAAAG) mutant, data are represented as mean ± SEM of triplicate measurements from at least 3 biological replicates. For the P13 mutants, data are represented as mean ± SEM of triplicate measurements from a single biological replicate.
Figure S4. Design and characterization of the trans aptamer-SL13 EMSA system, related to Figure 3. (A) Secondary structures of the isolated wild type env8HyCbl aptamer domain (nucleotides 1-81) and the regulatory stem loop (SL13, nucleotides 89-102). SL13 was 5' end labeled using $^{32}$P as described in materials and methods. (B) Representative ITC thermogram of HyCbl binding to the isolated wild type env8HyCbl aptamer domain with the $K_D$ for the interaction shown above the trace. Data is shown as mean ± SD of three independent titrations. (C) Representative gel shifts for each of the mutants depicted in Figure 3A, where the color of the boxes that outline each gel correspond to the quantified data. Dashed and solid outlines denote EMSAs conducted in the absence or presence of 10 µM HyCbl, respectively. (D) Representative gel shifts for each of the mutants depicted in Figure 3B, where the color of the boxes that outline each gel correspond to the quantified data. Dashed and solid outlines denote EMSAs conducted in the absence or presence of 10 µM HyCbl, respectively.
Figure S5. Cell density normalized GFPuv expression of riboswitches with altered linker sequence composition, related to Figure 4. (A) Mutations made to J1/13 with the sequence of the wild type linker shown at the top colored red. (B) Cell density normalized fluorescence of cells expressing the RNAs depicted in A grown in the absence (gray bars) or presence (red bars) of 5 µM HyCbl. Levels of fluorescence for cells expressing the wild type RNA are shown as striped bars and the non-functional L5(UGAAAG) mutant (ko) is shown at the far right. Data are represented as mean ± SEM of triplicate measurements from at least three biological replicates.
Figure S6. Cell density normalized fluorescence of cells expressing riboswitches with varying linker lengths, related to Figure 5. (A) Comparison of GFPuv expression from E. coli expressing linker length variants grown in the absence (gray bars) or presence (red bars) of 5 μM HyCbl. Levels of fluorescence from cells expressing the wild type RNA is shown at 7 linker nucleotides (striped bars). Longer lengths of J1/13 were made by addition of nucleotides to the 3’ end of the wild type linker, while shorter linkers were created by deletion from the 3’ end. Nucleotide addition to the 3’ end of J1/13 followed a repetitive pattern of the wild type sequence. Data are represented as mean ± SEM of triplicate measurements from at least three biological replicates. (B) Linear correlation between reporter expression at low intracellular cobalamin concentrations (gray bars, A) and mutant riboswitches with linker lengths between 5 and twenty-one nucleotides.
Figure S7. Freely diffusing single molecule burst titrations using a subset of linker length mutants and photobleaching of the wild type construct, related to Figure 6.
FRET efficiency histograms from free diffusion studies of riboswitches with (A) 7, (B) 25, (C) and 36 nucleotide linkers. Rainbow color scheme denotes increasing HyCbl concentrations from low (red) to high (violet). Histograms are fit to a sum of two Gaussians to determine the fraction of folded molecules as described previously (Holmstrom et al., 2014). (D) Plot of the fraction of folded molecules reveals that constructs with longer linkers are less docked, even under saturating conditions of ligand, suggesting that the decreased rate of interdomain loop-loop docking dictates the loss of efficient repression of reporter expression (see text for details). (E) Representative donor (green) and acceptor (red) fluorescence trajectory of the wild type construct. Single step photobleaching events (black arrow), were used to ensure that only individual surface immobilized molecules were considered for data analysis. (F) Representative FRET efficiency (black) of the wild type construct. Orange lines represent a model generated using a two state hidden-Markov model (McKinney et al., 2006). Fluorescence histograms (right) and the corresponding fits reveal the two state folding behavior.
**Supplemental Table 1.** Alignment of the L5-L13 KL interaction within the Cbl-II class of cobalamin riboswitches, related to Figure 2.

Underlined nucleotides in columns labeled “P5/L5 hairpin” and “P13/L13 hairpin” are predicted to form the helix of their respective stem loops. Predicted KL sequences are shown in the column at the far right, with bugled nucleotides shown in bold.

**Supplemental Table 2.** Sequences used in this study, related to Experimental Procedures.

*a* The wildtype *env8HyCbl* riboswitch sequence (underlined) is shown with the upstream promoter (italicized) and the linker region, J1/13, is highlighted by bold underline. For all other constructs, only the riboswitch sequence is shown with mutations made to the sequence denoted by bold underline.
Supplemental Experimental Procedures.

RNA Synthesis and Preparation

For all in vitro assays, templates for RNA transcription were amplified using PCR and transcribed by T7 RNA polymerase (Edwards et al., 2009). Transcription reactions were purified using the appropriate percentage denaturing polyacrylamide gel (8 M urea, 29:1 acrylamide:bisacrylamide) based on RNA length. Full length transcripts were visualized by UV shadowing, excised from the gel and removed by soaking at 4 °C in 0.5x TE buffer (10 mM Tris-HCl, pH 8.0, 1 mm EDTA). Buffer exchange and concentration were performed using centrifugal concentrators (Amicon) with the appropriate molecular weight cutoff. Final RNA concentration was calculated using the molar extinction coefficient as determined using an extinction coefficient calculated as the summation of the individual bases and the absorbance at 260 nm.

Preperation of smFRET Constructs

Riboswitch RNAs used for smFRET studies were generated via enzymatic ligation of commercially synthesized (IDT) custom oligonucleotides. First, an oligonucleotide containing an amino modified dT at position U24 (5′-biotin-AAAAAAAGGCCUAAAAGCGUAGUGGGAAA G[dT*]GACGUGAAAAUUCGUCCAGAUAC-3′) was covalently coupled to a mono-reactive Cy5 NHS ester (GE healthcare) following the manufacturer’s suggested protocol. The excess uncoupled and unreacted dye was removed via microcentrifuge gel filtration column (Thermo Scientific). This 5′ half of the env8HyCbl riboswitch was then enzymatically ligated to one of three oligonucleotides in order to generate each of the 3 different RNA constructs with variable J1/13 linker lengths (underlined sequences): 7 nucleotide linker (5′-phosphate-UUG AUACGGUUUAUCUCCGAAUGCCACCUCUAGGCCAUACAACGAGCAAGGAGACUC-Cy3-3′), 25 nucleotide linker (5′-phosphate-UUGAUACGGUUUAUCUCCGAAUGCCACCUCUAGGCCA
ACAACAUACAACAUACAACAUACAACAUACGAGCAAGGAGACUC-Cy3-3'), and 36 nucleotide linker (5’-phosphate-UUGAUACGGGUUUAUACUCCGAAUGCCACCUCUAGGCCAUACAACAUACAACAUACAGCAAGGAGACUC-amino-3’). Prior to ligation, the 36 nucleotide oligonucleotide was allowed to react with a mono-reactive Cy3NHS ester (GE healthcare) to covalently couple the oligonucleotide to the fluorophore. Standard non-splinted ligation was used to prepare the ligated 7 nucleotide and 25 nucleotide constructs as described previously (Holmstrom et al., 2014). The 36 nucleotide construct was generated via T4 RNA Ligase II (New England Biolabs, NEB) splinted ligation in T4 RNA Ligase reaction buffer. Prior to ligation, the 5’ half of the env8HyCbl riboswitch and the 36 nucleotide linker oligonucleotides were annealed to a 36 nucleotide (5’-GGAGTATAACCGTATCAAGTAATCTGGACGAATTTC-3’) DNA splint (IDT) by slowly cooling to room temperature from 85 °C. Following ligation, the DNA splint was removed via RNAse free DNAse I (NEB) digestion. Prior to experimentation, all smFRET constructs were purified via reverse phase ion-pair HPLC, buffer exchanged into storage buffer (125 mM KCl, 50 mM HEPES, 25 mM KOH, 0.1 mM EDTA, pH 7.5), and stored at -80 °C.

**Isothermal Titration Calorimetry (ITC)**

ITC was performed using RNAs that were synthesized and purified as described above. Following gel extraction and buffer exchange into 0.5X TE (10 mM Tris-HCl, pH 8.0, 1 mM EDTA), each RNA was dialyzed overnight at 4 °C into 1x ITC buffer (1 mM MgCl₂, 100 mM KCl, 10 mM NaCl, 50 mM K-HEPES, pH 8.0). Briefly, the desired volume of RNA was mixed with an equal volume of 1X ITC buffer and dispensed into 6-8000 Dalton molecular weight cutoff dialysis tubing (Spectra/Por). The RNA was dialyzed overnight in 1L of 1x ITC buffer at 4°C by gentle stirring. Appropriate concentrations of RNA for each titration were made by dilution of the dialyzed RNA into 1X ITC buffer. Stock solutions of HyCbl were made by dissolving solid HyCbl into 1X ITC buffer. For each titration, solutions of the desired HyCbl concentration were made
using absorbance maxima at 279, 359, and 537 nm (ε 19000, 20600, 9500 respectively). Titrations were performed at 37 °C using a MicroCal ITC$_{200}$ microcalorimeter (GE Healthcare) and data were fit using the Origin software suite as previously described (Gilbert and Batey, 2009).

**Structure Probing by Selective 2'-Hydroxyl Acylation Analyzed by Primer Extension (SHAPE)**

Chemical probing was performed using N-methylisatoic anhydride (NMIA) modification of the 2'-OH as previously described (Merino et al., 2005; Wilkinson et al., 2006), with slight modifications. For probing reactions, 10 µL reactions were prepared to final concentrations of 100 nM RNA, 100 mM Na-HEPES pH 8.0, 100 mM NaCl, 1 mM MgCl$_2$, 10 µM HyCbl, and 6.5 mM NMIA. Probing reactions were incubated at 37 °C for 41 minutes (5 half-lives). Following probing, primer extension using reverse transcriptase was carried out as previously described. Products were separated using 12% denaturing polyacrylamide gel electrophoresis and visualized using a Typhoon 9400 PhosphorImager (Molecular Dynamics). Reactivity profiles for each RNA were calculated using the SAFA software package (Das et al., 2005).

**Single-Molecule FRET experiments**

For surface immobilized experiments, biotin-streptavidin surface immobilization techniques were used to non-covalently attached individual FRET-labeled RNA constructs to a glass coverslip with a surface density of < 1 molecule per µm$^2$. Red-green, false-color, representations of surface immobilized fluorescent molecules were generated by scanning a 10 µm by 10 µm region of the coverslip over the objective using a piezo-electric scanning stage. The stage was then used to position an individual fluorescence feature, corresponding to a single FRET labeled
RNA construct, at the diffraction-limited focus of the 1.2 NA microscope objective. For data analysis, the photon arrival times were binned at 20 ms to generate fluorescence time-trajectories. The fluorescence signal was converted to FRET efficiency using the ratio of acceptor fluorescence to total fluorescence, before a simple thresholding-routine was used to determine the dwell times in the high and low FRET states. Histograms of the 100-1000 dwell times from 10-100 molecules was fit to a mono-exponential decay to determine the kinetic rate constants for docking and undocking. Although the low surface density of surface immobilized molecules greatly reduces the probability of having two molecules within one fluorescence feature, only those features that displayed single-step photobleaching were considered for data analysis to ensure that each time-trajectory was that of a single molecule with one donor and one acceptor, rather than species containing multiple FRET pairs (Figures S7E and S7F).

Free diffusion experiments were conducted on the same instrument as described above, with fluorescently labeled RNA free in solution at dilute concentrations (250 pM), rather than being immobilized. The focus of the microscope objective was placed 15 μM above the coverslip surface allowing for the detection of molecules in solution that transiently diffuse through the detection volume, resulting in short (1 ms) bursts of photons. The FRET efficiency was calculated for those bursts containing greater than 35 photons. Additionally, PIE was used to directly excite the acceptor thus ensuring that only molecules with active acceptor fluorophores were considered for analysis. The resulting FRET efficiencies are presented as a histogram (Figures S7A-S7C) to show the probability of observing bursts of a given transfer efficiency, and fit with a sum of two Gaussians to determine the faction of folded molecules (Figure S7D) as described previously (Holmstrom et al., 2014).
Supplemental References

Das, R., Laederach, A., Pearlman, S.M., Herschlag, D., and Altman, R.B. (2005). SAFA: semi-automated footprinting analysis software for high-throughput quantification of nucleic acid footprinting experiments. Rna 11, 344-354.

Edwards, A.L., Garst, A.D., and Batey, R.T. (2009). Determining structures of RNA aptamers and riboswitches by X-ray crystallography. Methods Mol Biol 535, 135-163.

Gilbert, S.D., and Batey, R.T. (2009). Monitoring RNA-ligand interactions using isothermal titration calorimetry. Methods Mol Biol 540, 97-114.

Holmstrom, E.D., Polaski, J.T., Batey, R.T., and Nesbitt, D.J. (2014). Single-molecule conformational dynamics of a biologically functional hydroxocobalamin riboswitch. J Am Chem Soc 136, 16832-16843.

McKinney, S.A., Joo, C., Ha, T. (2006). Analysis of single-molecule FRET trajectories using Hidden Markov Modeling. Biophys J 91, 1941-1951.

Merino, E.J., Wilkinson, K.A., Coughlan, J.L., and Weeks, K.M. (2005). RNA structure analysis at single nucleotide resolution by selective 2'-hydroxyl acylation and primer extension (SHAPE). J Am Chem Soc 127, 4223-4231.

Wilkinson, K.A., Merino, E.J., and Weeks, K.M. (2006). Selective 2'-hydroxyl acylation analyzed by primer extension (SHAPE): quantitative RNA structure analysis at single nucleotide resolution. Nat Protoc 1, 1610-1616.