Supplemental Information For:

A transient intermediate RNA structure underlies the regulatory function of the *E. coli thiB* TPP translational riboswitch.

Katherine E. Berman[1], Russell Steans[2], Laura M. Hertz[1] and Julius B. Lucks[3,4]*

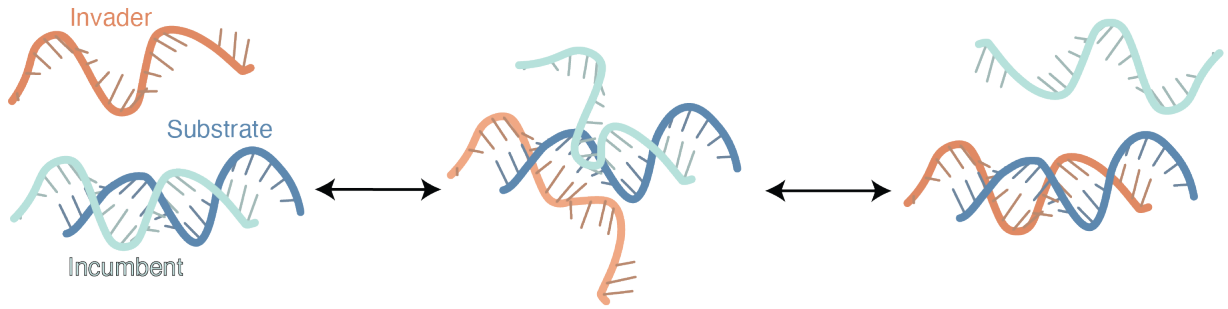
[1] Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL 60208

[2] Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208

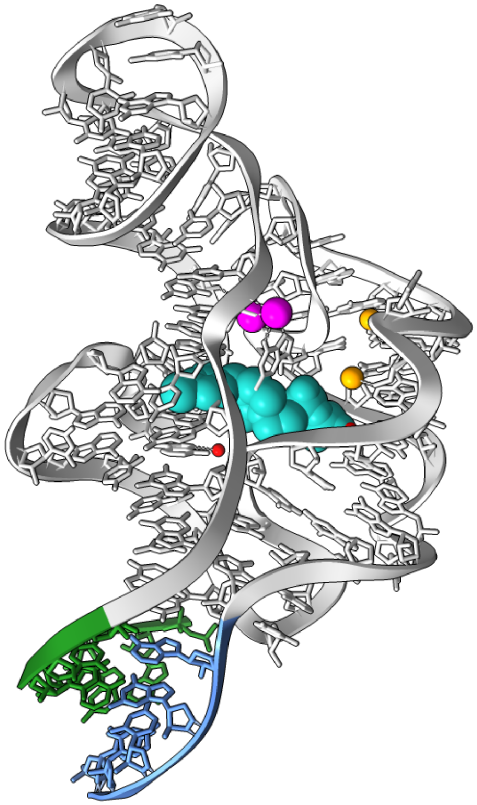
[3] Department of Chemical and Biological Engineering, Northwestern University, IL, 60208

[4] Center for Synthetic Biology, Northwestern University, Evanston, IL, 60208

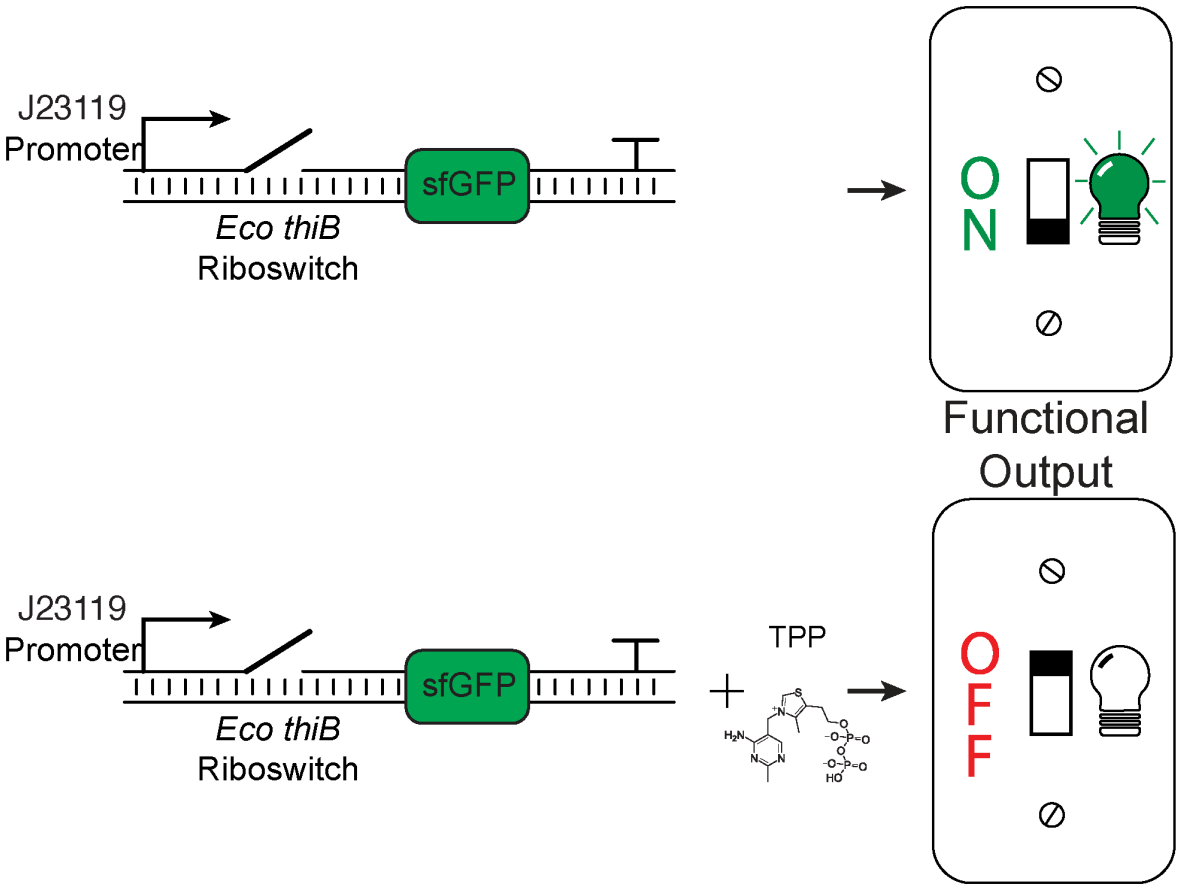
* To whom correspondence should be addressed: jblucks@northwestern.edu



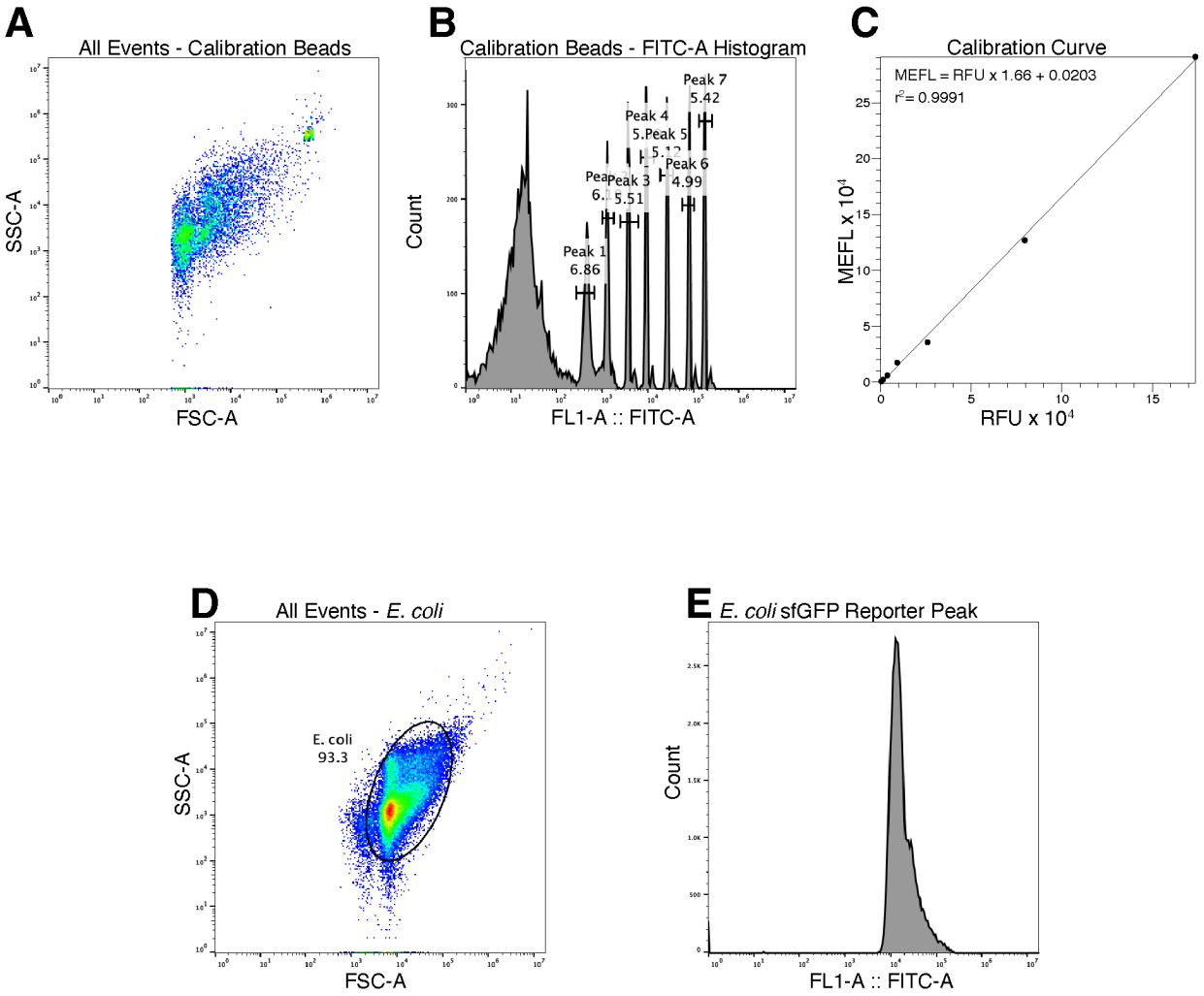
Supplementary Figure S1 | Schematic of nucleic acid strand displacement (Hong and Šulc 2019). The substrate (dark blue) and incumbent (light blue) strands are initially base paired. Pairing between the substrate and invader (orange) can displace the incumbent strand. If strand displacement is successful, the substrate and invader strand are completely base paired, releasing the incumbent.



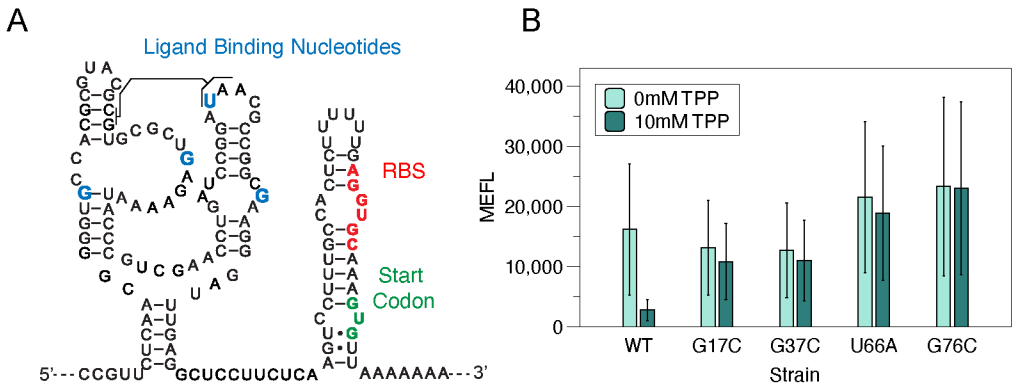
Supplementary Figure S2 | Crystal structure of *thiM* TPP Riboswitch.
Color-coded crystal structure representation of the *E. coli thiM* TPP riboswitch (PDB ID: 2GDI) (Serganov et al. 2006). The 5' side of the P1 helix is forest green, the 3' side of the P1 helix is light blue, and the ligand, thiamine pyrophosphate, is teal. Magnesium, potassium, and sodium ions are colored red, magenta and yellow respectively.



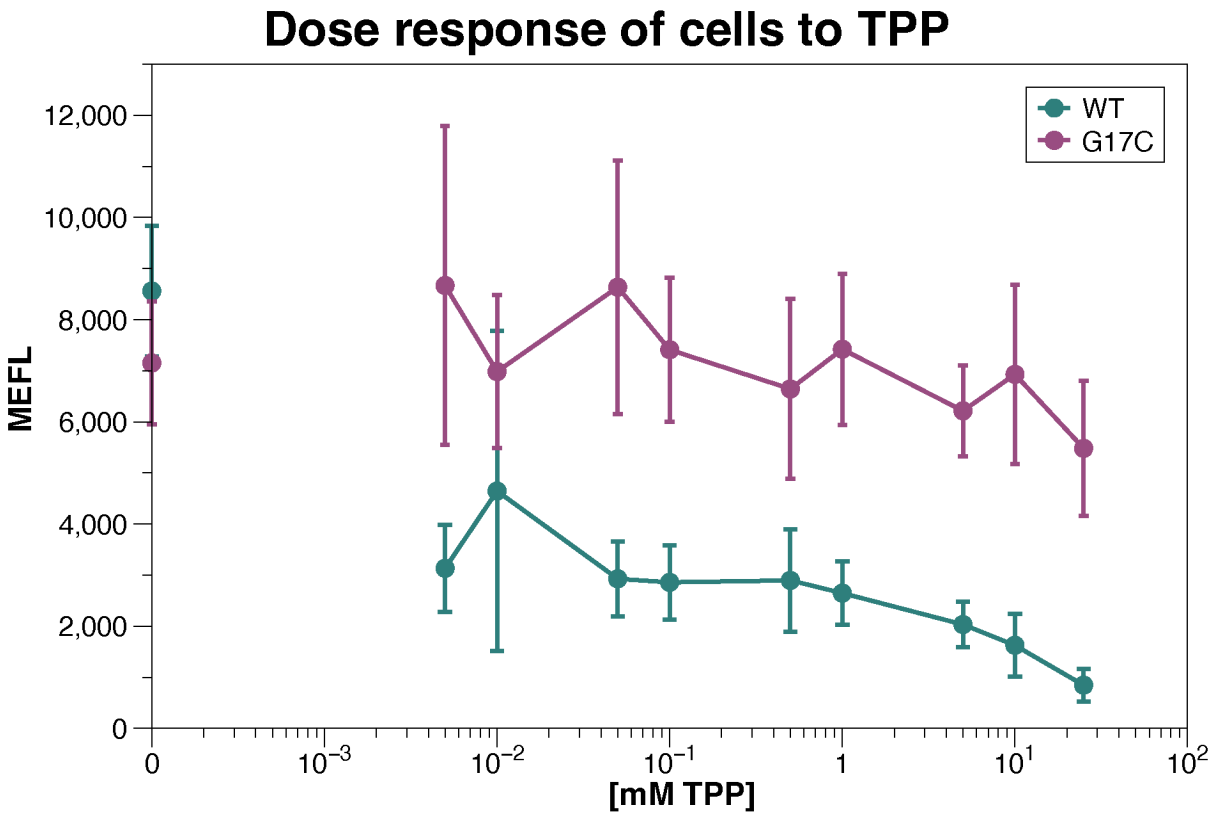
Supplementary Figure S3| *thiB* Riboswitch SFGFP Reporter Schematic. Schematic representing riboswitch regulated SFGFP expression plasmids. The bent arrow represents the J23119 constitutive *Escherichia coli* sigma 70 RNA polymerase promoter from the library of standardized parts; the switch symbol (broken top line) represents the *Escherichia coli thiB* TPP riboswitch sequence; the green colored block represents the SFGFP coding sequence; and the 'T' symbol represents the *Escherichia coli rrnB* terminator sequence. Sequences were cloned into a p15a plasmid backbone containing chloramphenicol antibiotic resistance. In the absence of thiamine pyrophosphate, the plasmid construct is expected to result in high amounts of SFGFP causing a fluorescent signal represented by the ON light switch, while in the presence of TPP, the plasmid construct is expected to result in low amounts of SFGFP, represented by the OFF light switch.



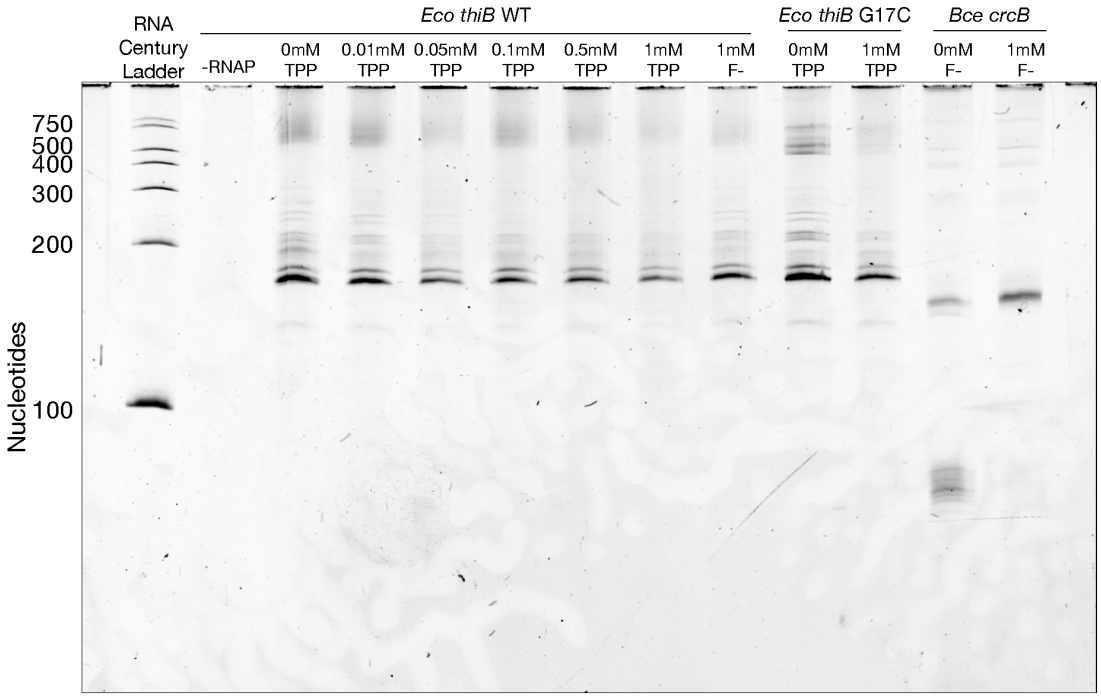
Supplementary Figure S4 | Flow Cytometry gating and calibration workflow. A) Representative calibration events collected for 8-peak rainbow Spherotech Calibration Beads (BD Biosciences) depicted in the FSC-A vs. SSC-A profile. For each calibration curve, 10,000 events were collected. B) The relative fluorescence units (RFU) of FITC measured by the BD Accuri C6 Plus flow cytometer for Spherotech Calibration 8-peak Rainbow Beads. Each relative fluorescent peak corresponds to a known mean equivalent fluorescein (MEFL) value supplied by the manufacturer. C) The geometric mean of each RFU calibration peak plotted against the known MEFL. The calibration curve was calculated using linear regression. The resulting linear regression equation was used to convert RFU values into MEFL units of all *E. coli* sfGFP reporter fluorescence values. D) All events collected for a sample of *E. coli* expressing sfGFP plotted as FSC vs. SSC from the flow cytometer. An ellipsoid gate was drawn around the highest density of events along this plot to select the population of *E. coli* cells. E) FITC histogram of RFU for the same *E. coli* sample. The histogram is within the *E. coli* gate depicted in D. Geometric means of RFU values were converted to MEFL using the calibration curve linear regression equation depicted in C. Fluorescent measurements were collected using BD C6 Accuri CSampler Software. Plots, Histograms, and RFU geometric means were created and calculated using FlowJo 10.8.0 software.



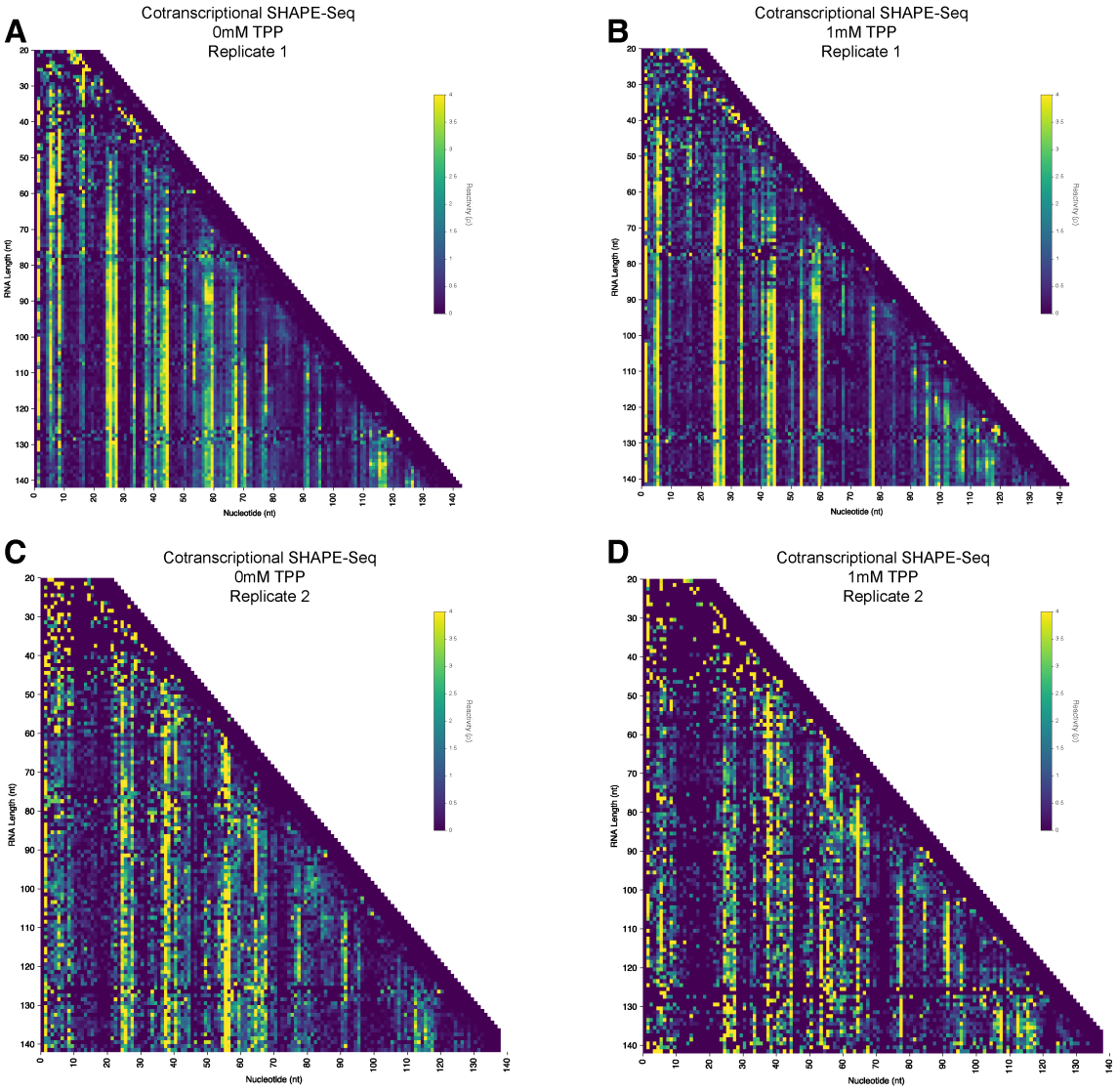
Supplementary Figure S5 | Ligand unresponsive mutant control assay. A) Secondary structure drawing of the *Eco thiB* TPP riboswitch in the translational 'off' structure with ribosome binding site (red), start codon (green) and mutated nucleotides in ligand unresponsive mutants (blue) highlighted. B) Flow cytometry assay of ligand unresponsive mutants with and without the presence of 10mM TPP supplied to the media. Data in B represent averages over 3 biological replicates, each performed in three technical replicates for a total of nine datapoints (n=9). Error bars representing standard deviation.



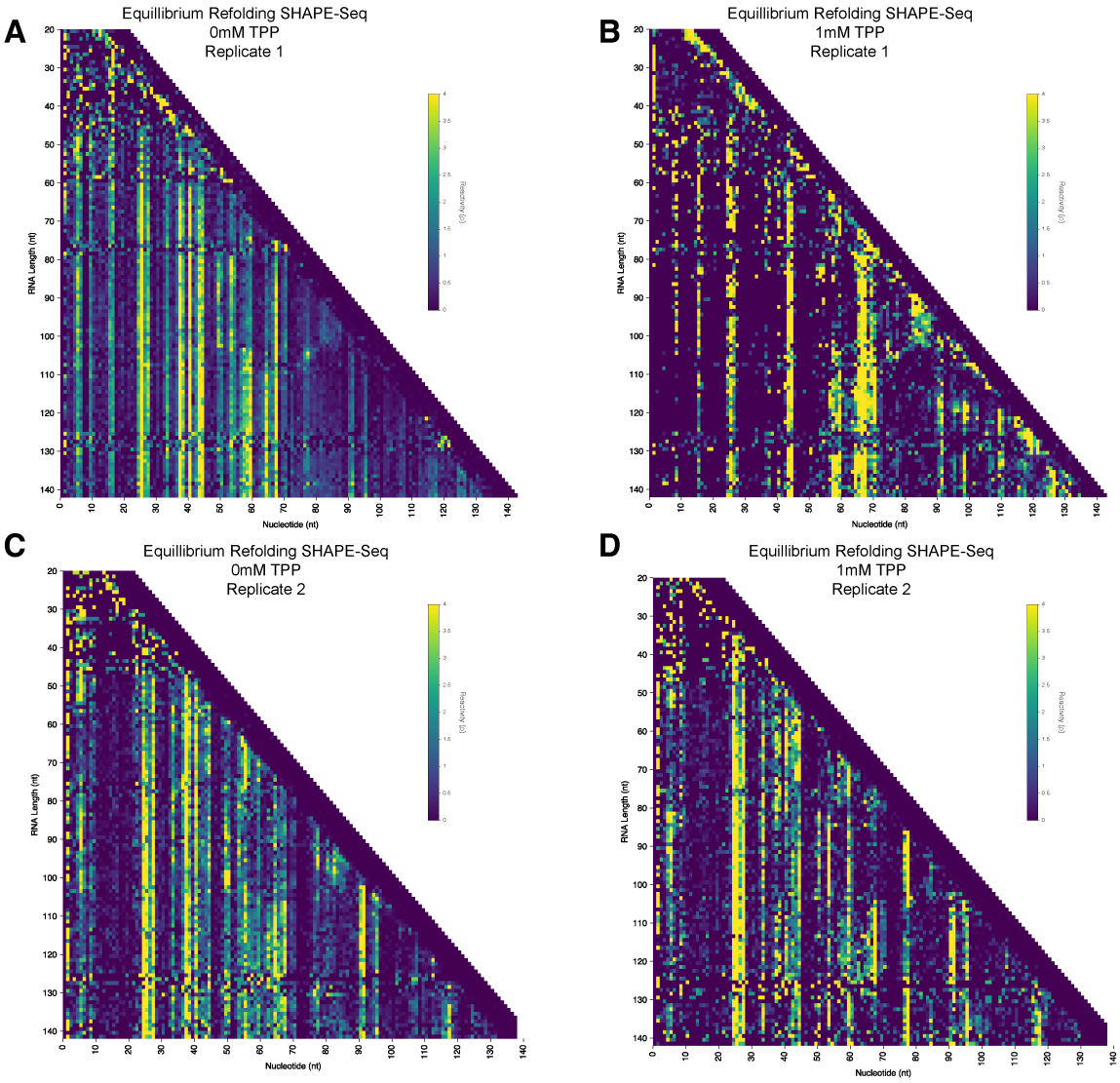
Supplementary Figure S6 | Dose response curve of riboswitch constructs in response to externally supplied TPP. Measurements were observed for *E. coli* cells expressing TPP riboswitch regulated GFP at 10 different concentrations of TPP. SFGFP fluorescence was measured using flow cytometry after 6 hours of growth in M9 media. Data represented is 3 triplicates collected on three different days for nine datapoints (n=9), with error bars representing standard deviation. WT represents the wildtype *E. coli thiB* TPP riboswitch sequence and G17C represents a single nucleotide mutant ligand unresponsive control. 0mM TPP points are plotted on the y-axis. With increasing concentration of TPP, sfGFP expression decreases.



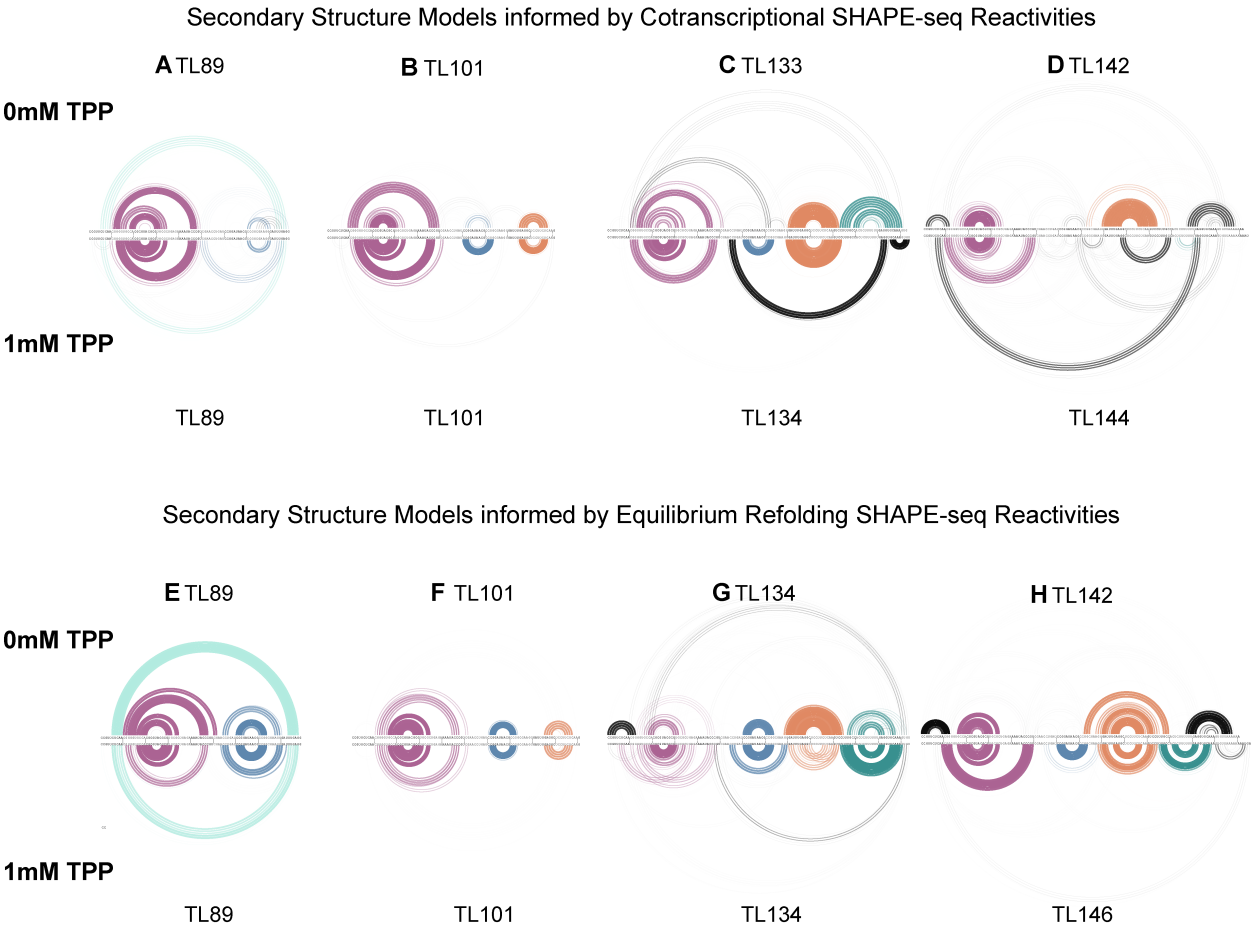
Supplementary Figure S7 | Replicate single round *E. coli* RNAP *in vitro* transcription assay. Technical replicate of Figure 1C. Single round *E. coli* RNAP *in vitro* transcription assay measuring transcription product lengths on 10% Urea polyacrylamide gel electrophoresis. Predicted full length transcript product of *Eco thiB* is 173nt. Predicted full length transcript product of *Bce crcB* is 124nt and the terminated transcript product is 82nt.



Supplementary Figure S8 | Cotranscriptional SHAPE-seq Matrices of *Eco thiB* TPP Riboswitch . Cotranscriptional SHAPE-seq data for all intermediate transcript lengths of the *Eco thiB* TPP riboswitch in the absence (A,C) and presence (B,D) of TPP. High reactivity (yellow) corresponds to nucleotide flexibility and low reactivity (blue) corresponds constrained nucleotide structure such as base pairing. Reactivity is calculated and reported as rho reactivity (Loughrey et al. 2014).

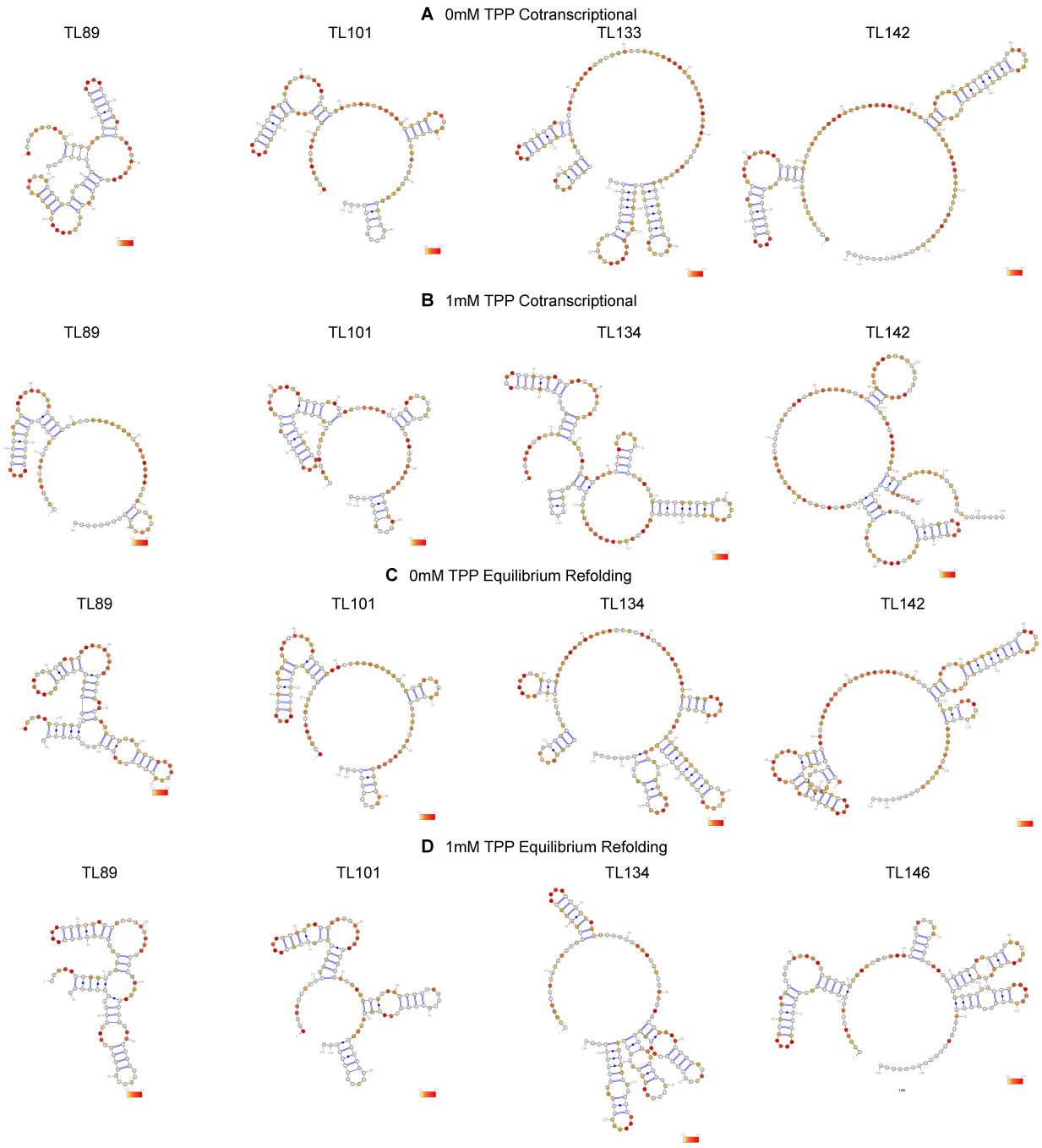


Supplementary Figure S9 | Equilibrium Refolding SHAPE-seq Matrices of *Eco thiB* TPP Riboswitch. Equilibrium Refolding SHAPE-seq data for all intermediate transcript lengths of the *Eco thiB* TPP riboswitch in the absence (A,C) and presence (B,D) of TPP. High reactivity (yellow) corresponds to nucleotide flexibility and low reactivity (blue) corresponds constrained nucleotide structure such as base pairing. Reactivity is calculated and reported as rho reactivity (Loughrey et al. 2014).



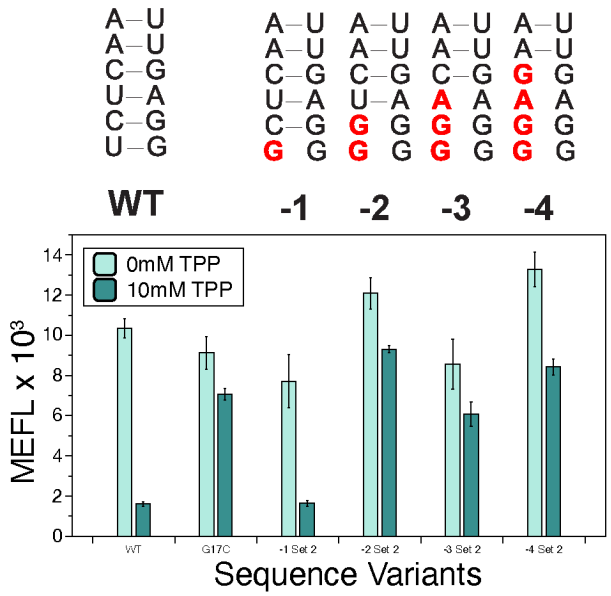
Supplementary Figure S10 | RNAbow plots demonstrating secondary structure models informed by SHAPE-seq Reactivities Replicate 2. Secondary structure models generated by R2D2 (Yu et al. 2021) using replicate 2 cotranscriptional and equilibrium refolded SHAPE-seq data for the *thiB* TPP riboswitch both with and without ligand. SHAPE-seq experiments were performed at all transcript lengths, with representative lengths shown here. R2D2 produced 100 secondary structure models of the intermediate RNA folds for all transcript lengths that are maximally consistent with reactivity data. RNAbow plots (Aalberts and Jannen 2013) represent base pairing with an arc between nucleotides, and the opacity of the arc indicates the prevalence of the structure in the one hundred modelled pathways. Arcs are colored based on the secondary structure of the riboswitch, with light teal referring to P1, purple referring to P2/P3, blue referring to P4/P5, terracotta referring to the anti-sequestering stem and dark teal referring to the sequestering stem. Representative lengths were chosen to be similar but do not exactly match due to differences in read coverage (see Methods).

Consensus Secondary Structure Models informed by SHAPE-seq Data Replicate 2

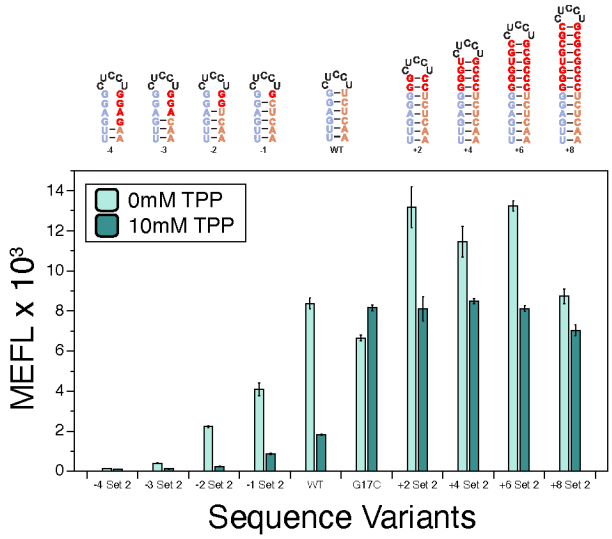


Supplementary Figure S11 | Consensus secondary structures demonstrating models informed by SHAPE-seq Reactivities Replicate 2 drawn using VARNA. Consensus secondary structure models generated by R2D2 using replicate 2 cotranscriptional and equilibrium refolded SHAPE-seq data for the *thiB* TPP riboswitch both with and without ligand. SHAPE-seq experiments were performed at all transcript lengths, with representative lengths shown here. R2D2 produced a secondary structure model of the RNA folding pathway for all transcript lengths 100 times that are maximally consistent with reactivity data. Secondary

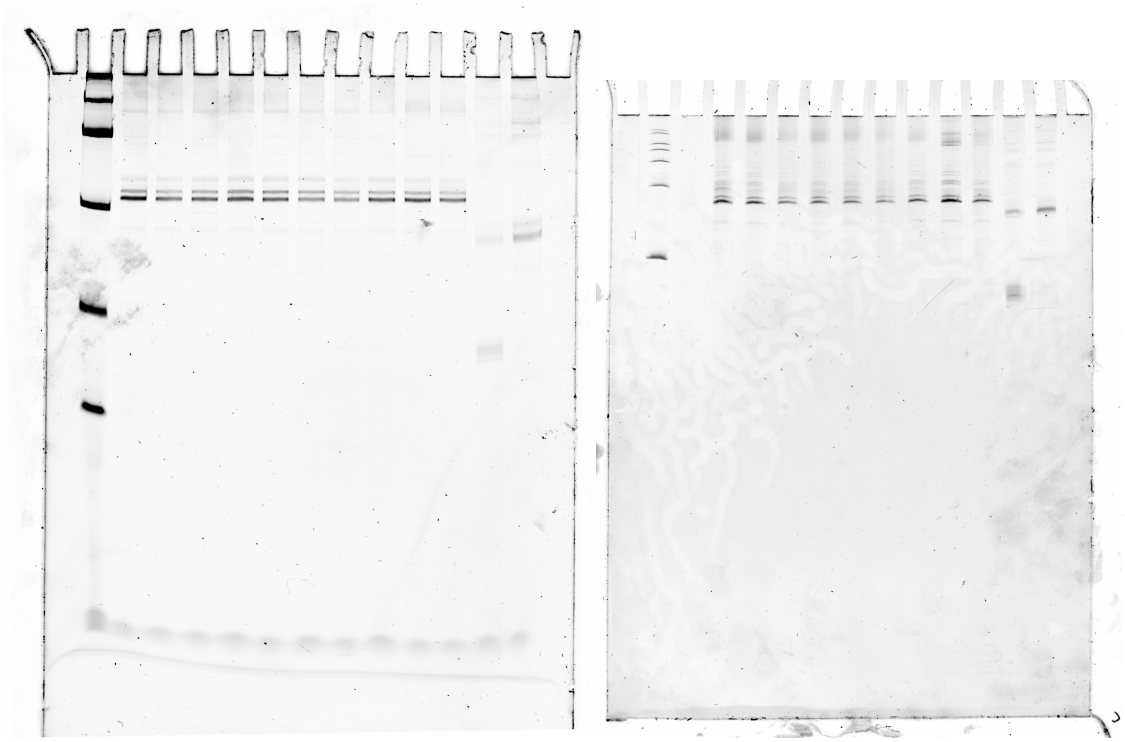
structures are drawn using VARNA (Darty et al. 2009) from the consensus paired regions within the 100 modelled RNA folding pathways. Red coloring refers to how unstructured nucleotides are predicted to be, based on secondary structure models, with red being more unstructured.



Supplementary Figure S12 | *In vivo* TPP response of P1 mismatch mutants. Second set of mismatch mutations were designed to shorten the P1 hairpin. Mutants have a similar phenotype to Figure 5A. Flow cytometry measurements of sfGFP fluorescence in *E. coli* cells were taken after 6 hours of growth in M9 media. Bar graphs represent mean values across one biological replicate, each performed in technical triplicate for three total datapoints (n=3). Error bars represent the standard deviation from the mean. WT represents the wildtype *E. coli thiB* TPP riboswitch sequence and G17C represents a single nucleotide mutant ligand unresponsive control.



Supplementary Figure S13 | *In vivo* TPP Response of a second set of anti-sequestering stem mutants. Second set of anti-sequestering stem shortening and lengthening mutants that have a similar phenotype to mutants in Figure 5B. Flow cytometry measurements of sfGFP fluorescence in *E. coli* cells were taken after 6 hours of growth in M9 media. Bar graphs represent mean values across one biological replicate, each performed in technical triplicate for three total datapoints (n=3). Error bars represent the standard deviation from the mean. WT represents the wildtype *E. coli thiB* TPP riboswitch sequence and G17C represents a single nucleotide mutant ligand unresponsive control.



Supplementary Figure S14 | Raw Urea-PAGE Images. Unedited and raw images of urea polyacrylamide gel images from main text Figures 1C (left) and S7 (right).

Supplemental Table S1 | Template sequences and primers for single round *in vitro* transcription and SHAPE-seq. DNA template sequences were also used to generate biotinylated templates for cotranscriptional and equilibrium refolding SHAPE-Seq experiments.

Description	Sequence
Forward Template Primer	ataagcttccgatggcgcgc
Reverse Template Primer	/5Biosg/ctccagtgaaaagttcttctcctttgctc
E. coli thiB TPP Riboswitch WT Template	ataagcttccgatggcgcgccgagaggctttacactttatgcttccggctgaattctaaagatctttgaca gctagctcagtcctaggtataatactagtccgttctcaacggggtgccacgcgtacgcgtgcgctgaga aaatacccgtcgaacctgatccggataacgccggcgaagggatttgaggctccttctcaagtcctttgc cactctttttgaggtgcaaagtgttaaaaaaaatgagcaaaggagaagaacttttcactggag
E. coli thiB TPP Riboswitch G17C Template	ataagcttccgatggcgcgccgagaggctttacactttatgcttccggctgaattctaaagatctttgaca gctagctcagtcctaggtataatactagtccgttctcaacggggtcccacgcgtacgcgtgcgctgaga aaatacccgtcgaacctgatccggataacgccggcgaagggatttgaggctccttctcaagtcctttgc cactctttttgaggtgcaaagtgttaaaaaaaatgagcaaaggagaagaacttttcactggag
Bce crcB F-Riboswitch WT Template	Ataagcttccgatggcgcgccgagaggctttacactttatgcttccggcttgattctaaagatctttgaca gctagctcagtcctaggtataatactagtttataggcgatggagttcgccataaacgctgcttagctaatg actcctaccagtatcactactggtaggagtctatttttttaggaggaaggatctatgagcaaaggagaa gaacttttcactggag
RNA Linker	/5Phos/CUGACUCGGGCACCAAGGA/3ddC/
Reverse Transcription Primer	GTCCTTGGTGCCCGAGT
DNA Adaptor/dumbbell ligation primer	/5Phos/TGAAGAGCCTAGTCGCTTTGCANNNNNNCTGCCC/3SpC3/
Selection Primer (+)	CTTTCCCTACACGACGCTCTTCCGATCTRRRYGTCCTTGGTGCCCGA G*T*C*A*G
Selection Primer (-)	CTTTCCCTACACGACGCTCTTCCGATCTYYYRGTCCTTGGTGCCCGA G*T*C*A*G
PE_F*	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGC TCTTCCGATCT
Illumina Reverse Primers (TruSeq)*	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGACTGGAGTTCAGA CGTGTGCTCTTCCGATCTTGAACAGCGACTAGGCTCTTCA

*Oligonucleotide sequences ©2007-2013 Illumina, Inc. All rights reserved.

Supplementary Table S2 | Small Read Archive (SRA) deposition table.
Sequencing data generated in this paper has been made available on the Small Read Archive (https://www.ncbi.nlm.nih.gov/sra). This data is available via BioProject PRJNA950133 or using the individual accession numbers below:

SRA Accession	Experiment Description	Figures
SAMN33977272	Cotranscriptional SHAPE-seq 0mM TPP Replicate 1	Figs 5 & S10
SAMN33977273	Cotranscriptional SHAPE-seq 0mM TPP Replicate 2	Figs S10,S12
SAMN33977271	Cotranscriptional SHAPE-seq 1mM TPP Replicate 1	Figs 5 & S10
SAMN33977274	Cotranscriptional SHAPE-seq 1mM TPP Replicate 2	Figs S10,S12
SAMN33977270	Equilibrium Refolding SHAPE-seq 0mM TPP Replicate 1	Figs 5 & S11
SAMN33977275	Equilibrium Refolding SHAPE-seq 0mM TPP Replicate 2	Figs S11,S12
SAMN33977269	Equilibrium Refolding SHAPE-seq 1mM TPP Replicate 1	Figs 5 & S11
SAMN33977276	Equilibrium Refolding SHAPE-seq 1mM TPP Replicate 2	Figs S11,S12

Supplementary Note S1 | SHAPE-seq Data Analysis.

Spats:

Download and Installation:

Software for Spats can be downloaded from GitHub:
http://luckslab.github.io/spats/index.html

Detailed installation and documentation for the use of Spats can be found here:
https://spats.readthedocs.io/en/master/

Usage:

The following parameters were used in the spats.config files to run Spats to calculate reactivities for each transcript length in SHAPE-seq experiments.

ignore_stops_with_mismatched_overlap=False
allow_indeterminate = False
minimum_target_match_length=10
allow_multiple_rt_starts=True
count_mutations = False
allowed_target_errors = 2
count_left_prefixes = True
collapse_left_prefixes = True
allowed_dumbbell_errors = 3
dumbbell = TGAACAGCGACTAGGCTCTTCA
cotrans = True

Alignment was completed to an .fa file containing the following sequence:

>thiB_fulllength_cotrans
CCGTTCTCAACGGGGTGCCACGCGTACGCGTGCGCTGAGAAAATACCCGTCGAACCTGAT
CCGGATAACGCCGGCGAAGGGATTTGAGGCTCCTTCTCAAGTCCTTTGCCACTCTTTTTTG
AGGTGCAAAGTGTTAAAAAAAATGAGCAAAGGAGAAGAACT

R2D2:

Download and Installation:
R2D2 can be downloaded from GitHub to a Linux server:
https://github.com/LucksLab/R2D2

Usage:
Input files are SHAPE-Seq reactivities generated by Spats 2.0.5 using the command:
spats_tool dump old_txt

The following parameters were used for cotranscriptional SHAPE-Seq data analyzed by Spats reported in this work: python <installation_dir>/R2D2/analyze_cotrans_SHAPESeq.

py --in_dir <reactivity_dir> --out_dir <output_dir> --adapter
python $code_dir/analyze_cotrans_SHAPE-Seq.py --in_dir $data_dir --out_dir $out_dir --adapter "CTGACTCGGGCACCAAGGA" --e 50000 --endcut 0 --constrained_c "3.5" --scale_rho_max "1" --draw_all "True" --most_count_tie_break "False" --weight_paired "0.8" --scaling_func "K" --cap_rhos "True" --pol_fp "14" --p 1

The following parameters were used for equilibrium-refolded SHAPE-Seq data analyzed by Spats reported in this work:
python $code_dir/analyze_cotrans_SHAPE-Seq.py --in_dir $data_dir --out_dir $out_dir --adapter "CTGACTCGGGCACCAAGGA" --e 50000 --endcut 0 --constrained_c "3.5" --scale_rho_max "1" --draw_all "True" --most_count_tie_break "False" --weight_paired "0.8" --scaling_func "K" --cap_rhos "True" --pol_fp "0" --p 1

Supporting Data File Descriptions

Berman_thiB_cotrans_Supp_All_Source_Data.xlsx – Calibrated flow cytometry data for all figures in this study.

Berman_thiB_Cotrans_Supp_DataFile_Sequences.xlsx – All plasmid sequences used in this study.

References:

Aalberts DP, Jannen WK. 2013. Visualizing RNA base-pairing probabilities with RNAbow diagrams. http://www.rnajournal.org/cgi/doi/10.1261/rna.033365.112.

Darty K, Denise A, Ponty Y. 2009. VARNA: Interactive drawing and editing of the RNA secondary structure. *Bioinformatics* 25: 1974–1975.

Hong F, Šulc P. 2019. An emergent understanding of strand displacement in RNA biology. *Journal of Structural Biology* 207: 241–249.

Loughrey D, Watters KE, Settle AH, Lucks JB. 2014. SHAPE-Seq 2.0: systematic optimization and extension of high-throughput chemical probing of RNA secondary structure with next generation sequencing. *Nucleic Acids Res* 42: e165–e165.

Serganov A, Polonskaia A, Phan AT, Breaker RR, Patel DJ. 2006. Structural basis for gene regulation by a thiamine pyrophosphate-sensing riboswitch. *Nature* 441: 1167–1171. http://www.nature.com/articles/nature04740.

Yu AM, Gasper PM, Cheng L, Lai LB, Kaur S, Gopalan V, Chen AA, Lucks JB. 2021. Computationally reconstructing cotranscriptional RNA folding from experimental data reveals rearrangement of non-native folding intermediates. *Molecular Cell* 81: 870-883.e10.