Supplementary Data

Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base-pairing

Sanshu Li\textsuperscript{1,2} and Ronald R. Breaker\textsuperscript{1,2,3, *}

\textsuperscript{1}Howard Hughes Medical Institute, \textsuperscript{2}Department of Molecular, Cellular and Developmental Biology, and \textsuperscript{3}Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

\textsuperscript{*}To whom correspondence should be addressed. Tel: +1 203 432 9389; Fax: +1 203 432 6161; Email: ronald.breaker@yale.edu.
**Supplementary Figure S1.** Common mechanisms of bacterial TPP riboswitch-mediated regulation (15). (A) Control of transcription elongation by ligand-mediated formation of an intrinsic transcription terminator. Terminator stems typically include a strong base-paired stem followed by a run of U nucleotides. The anti-terminator stem competes with the terminator stem in a ligand-dependent fashion to control transcription. (B) Control of translation initiation by ligand-mediated sequestration of the ribosome binding site (RBS). Ligand binding controls formation of an anti-anti-RBS stem, which determines whether the RBS becomes sequestered in a stem structure.
Supplementary Figure S2. Known mechanisms for eukaryotic riboswitch regulation (22,23,24,25,26). (A) In many fungal *NMT1* genes, sequestration of a 5′ splice site of an intron in the 5′ UTR occurs via base-pairing to the P4-P5 region of the TPP aptamer when it is not occupied by ligand. High TPP concentrations expose the alternative 5′ splice site to yield a spliced mRNA that carries uORFs which, when expressed, preclude expression of the main ORF. (B) In an intron of the 3′ UTR of many plant *THIC* genes, the mechanism of alternative splicing control is similar to that described in A. The diamond symbol represents the transcript processing site (TPS) whose removal in the presence of high TPP concentrations prevents processing and expression of the mRNA. (C) In the algal *THIC* gene, low TPP concentrations presumably cause alternative base-pairing (black rectangles) of the 5′ side of the P1 stem to sequester the first 3′ splice site. This favors alternative splicing at the second 3′ splice site located downstream of the first, to produce a short product. When the TPP concentration is high, the aptamer is structured and the first 3′ acceptor site is available for splicing, therefore producing longer spliced products. The dashed lines designate splicing between the various splice sites.
**Supplementary Figure S3.** Mutations to nucleotides located immediately upstream of 5′ splice site S4 do not greatly affect gene expression. The region between nucleotides 16 and 20 of construct WT* (see Supplemental Fig. S5) was subjected to single or double mutations designated by the nucleotides depicted in red. Splice site S4 (circled) and flanking regions depicted in blue are complementary to the blue nucleotides in the aptamer P4 and P5 region. This base-pairing potential could lead to structures that are similar to those observed in other fungal or plant TPP riboswitches wherein alternative base-pairing of the P4-P5 nucleotides with nucleotides flanking a proximal 5′ splice site controls alternative splice site selection. However, mutations that disrupt this possible alternative base-paired structure do not dramatically alter gene expression, and yield only ~2-fold differences or less in the fold modulation of gene expression. Fold modulation values were calculated as the expression level (RLU) in the absence of thiamin divided by the expression level in the presence of thiamin. Values are the average ± standard deviation from three independent replicates.
Supplementary Figure S4. Sequence, structure and gene expression activity of the natural NCU01977 riboswitch-reporter fusion construct carrying intron 1, intron 2 and either the full-length aptamer P3 stem (Full-length WT) or a truncated version of P3 (Full-length WT*) as depicted. Fold modulation of gene expression in the absence versus the presence of 30 μM thiamin supplementation using assay conditions as described in the main text. In data not shown here, the elimination of intron 1 reduces the effect of thiamin on gene control only by ~2 fold.
Supplementary Figure S5. In-line probing of TPP aptamers carrying either (A) the full-length (WT) or (B) the truncated (WT*) P3 stem. NR, T1 and "OH designate no reaction, partial RNA digestion with RNase T1, or partial RNA digestion with alkali, respectively. Bands corresponding to cleavage after certain G residues are numbered. Site 1 and site 2 show reduced cleavage when the aptamer is bound by TPP.
Supplementary Figure S6. Conserved α and α′ base-pairing using nucleotides of the right shoulder of the TPP aptamer P1 stem. Red rectangles designate potential base-pairing between α and α′. Blue shading identifies base-pairing regions of the aptamer P1 stem. Orange nucleotides conform to the consensus sequences of α and α′ present in 17 of 35 fungal species examined. The numbers are nucleotide positions relative to the start codon of the NCU01977 coding region according to available annotations, where negative numbers indicate positions upstream of the start codon. The riboswitch aptamer begins with the central blue-shaded nucleotides and ends with the blue-shaded nucleotides in the α′ element. The primary 5′ and 3′ splice sites are represented by the GU and AG nucleotides at the start and end of the sequences depicted. Organism names are abbreviated as in Fig. 4.
Supplementary Figure S7. Distances between key features of representative NCU01977 TPP riboswitches. (A) through (D) plot the distances between the features noted.
NCU01977 Gene Annotations (as of August 2009)

Supplementary Figure S8. NCU01977 gene sequence and annotations (according to the Neurospora crassa database published (44) and subsequently updated as of 2009 by the Broad Institute. Putative coding regions are shaded and all the splice site sequences GT and AG are depicted in red. Note that the putative coding exon encompassing nts 719-802 (84 nts) was not observed in our RT-PCR products and is unlikely to exist since it overlaps with the riboswitch domain (underlined). Alternative 5′ splice sites S1 (nts 246-247), S2 (nts 467-468), S3 (nts 495-496) and S4 (nts 613-614) are depicted in red. The binding sites for the forward primer (nts 504-
525) and the reverse primer (nts 963-982) used previously (23) are depicted in blue. The binding sites for the forward primer (nts 1-17) and the reverse primer (nts 1251-1274) used in the current study are depicted in yellow.
Additional NCU01977 Gene Annotations (as of July 2012)

CTGGAGTCTGTTTGATATTCAAGCAAATCGTACCGAAACATG
CTGGGCACCATCATCATCATCATCATGATC(-183 to -1)
ATGACAAACACACACCCGTTTTTTAGCCAGGTCAGGCCCAGCGTACGGGGAGCCAGC
CTGACTTGTCGCCGAGGATC (-183 to -1)

Supplementary Figure S9. NCU01977 gene sequence and annotations (according to the Neurospora crassa database available as of July 2012). The sequence from -183 to -1 is predicted to include another putative coding exon (gray shading) with a possible start codon (red) and a small intron.
Table S1. Sequences of DNA primers and methods to make plasmid constructs.

| Cloning of pLL07-2-1 |  |
|----------------------|--------------------------|
| DNA1 5′-CCCCGGGTCTAGCATCTAGAGGACGCCAAGAAC | pLL07-F (XbaI) |
| DNA2 5′-GTTCTTGCCGTCCTCTAGCTAGAGCCAAGCGG | pLL07-R (XbaI) |
| DNA3 5′-GATCCAATTGCCAGCGGGGCCTCCACCCTTC | Tub-F (MfeI) |
| DNA4 5′-GATCGAATTTCGACGGTTTGATGACGAACACGGG | Tub-R (EcoRI) |

As described previously (23), plasmid pLL07 (31), which carries a luciferase reporter gene, was mutated to disrupt the luciferase start codon and to insert an XbaI restriction site by using a QuickChange XL Site-Directed Mutagenesis Kit (Stratagene) with primers pLL07-F and pLL07-R. The promoter for the β-tubulin gene in *N. crassa* was amplified by PCR using primers Tub-F and Tub-R and inserted between MfeI and EcoRI in front of the luciferase (LUC) gene to obtain the plasmid pLL07-2-1.

Plasmid with wild-type TPP riboswitch (pTPP-WT)

|  |  |
|----------------------|--------------------------|
| DNA5 5′-ATGCGAATTCTGCTGTGAGATAAACACCAAAAC | VivoTPP-F (EcoRI) |
| DNA6 5′-ATCGTCTAGACCCGCAACGGTG | VivoTPP-R (XbaI) |

To construct a plasmid carrying the wild-type TPP riboswitch, 100 ng *Neurospora crassa* genomic DNA (gDNA), primers VivoTPP-F and VivoTPP-R and Phusion High-Fidelity DNA Polymerase (New England BioLabs) were used for PCR. The resulting PCR fragment was digested with EcoRI and XbaI and ligated into pLL07-2-1, which had been digested by EcoRI and XbaI and purified. The resulting plasmid was called pTPP-WT.

TPP riboswitch plasmid without the first intron (pMSWT)

|  |  |
|----------------------|--------------------------|
| DNA7 5′-CTATTCCGCTCGCTTTGAAC | No-first-intron-F |
| DNA8 5′-CAAAGCGGAGCGAATACCCGCTGTGTG | No-first-intron-R |

Two-step PCR was used to make the plasmid containing the TPP riboswitch lacking the first intron. Plasmid pTPP-WT was used as a template. Primers VivoTPP-F and No-first-intron-R were used to make fragment I and primers No-first-intron-F and VivoTPP-R were used to make fragment II. Then these two pieces were joined by PCR using primers VivoTPP-F and VivoTPP-R. The ligated fragment was digested with EcoRI and XbaI and cloned into pLL07-2-1. The resulting plasmid was called plasmid pMSWT.

Construction of splice site mutants

|  |  |
|----------------------|--------------------------|
| DNA9 5′-CTTCATCGCTTCGGAGATGAAAGCACA | Sp1-F |
| DNA10 5′-GTGTGCTTTCCACTCTCCGAAGCGATGAAG | Sp1-R |
| DNA11 5′-CGGCGATGACATAGGAATAAGTTTAGCGACAGAC | Sp2-F |
| DNA12 5′-GTCTGCTCAACTTATCTCTATGTCATCGCCG | Sp2-R |
| DNA13 5′-GACTGACAGCATAAGGATCGTTCGCAACCAAT | Sp3-F |
| DNA14 5′-GATTGCTGCGCAAGCATACTATGTCATGTC | Sp3-R |
| DNA15 5′-CATGGATATTACTCAAGGATAGGAGGTTGC | Sp4-F |
Similarly, two-step PCR was used to make fragments for splice site mutants using plasmid pMSWT as a template. For example, to construct splice site 1 mutant Sp1, primers VivoTPP-F and Sp1-R were used to make fragment I and Sp1-F and VivoTPP-R were used to make fragment II. Next, fragments I and II were joined by PCR using primers VivoTPP-F and VivoTPP-R. The ligated fragment was digested with EcoRI and XbaI and cloned into pLL07-2-1. The resulting plasmid was transformed into Neurospora crassa to create mutant Sp1. Similarly, the Sp2 mutant was made by primers Sp2-F and Sp2-R, the Sp3 mutant by Sp3-F and Sp3-R and the Sp4 mutant by Sp4-F and Sp4-R.

| TPP riboswitch plasmid with shortened P3 stem and without the first intron (pNoIP) |
|---------------------------------|---------------------------------|
| DNA16  5'-GCAACCTCTCATCTATCTTGTAGTAATATTCATG | Sp4-R  |

Two-step PCR was used to make fragments for this construct using plasmid pMSWT as a template. Primers VivoTPP-F and Shortened-P3-stem-R were used to make fragment I and primers Shortened-P3-stem-F and VivoTPP-R were used to make fragment II. These two pieces were joined by PCR using primers VivoTPP-F and VivoTPP-R. The ligated fragment was digested with EcoRI and XbaI and cloned into pLL07-2-1. The resulting plasmid was called plasmid pNoIP.

Construction of serial deletion mutants

| DNA19  5'-GAGAGGGTGCGCGCTCACTAC | Del 1-F |
| DNA20  5'-GTGAGCGCAACCTCTCCTCCTCGCCGGCGATGCTTTTG | Del 1-R |
| DNA21  5'-GTGAGCGCAACCTCTCCTCCTCCTCTAGCTTGCTACTTCTCATG | Del 2-R |
| DNA22  5'-GTGAGCGCAACCTCTCCTCCTCCTACTACCGCGAAG | Del 3-R |
| DNA23  5'-GAGGCCATGATATTACTCAAG | Del 4-F |
| DNA24  5'-GAGTAATATCCATGATCCCGCTCTGATCTTTCATG | Del 4-R |
| DNA25  5'-GAGTAATATCCATGATCCCGCTCTGATCTTTCGTTG | Del 5-R |
| DNA26  5'-TGCCACCGCTCTCGAGG | Del 6-F |
| DNA27  5'-CTCGAGAGAAGGTGCCAGCTACCGCGTCTCTGATCTTTCATG | Del 6-R |
| DNA28  5'-AGTAATATCCATGATCCCGCTCTGATCTTTCGTTG | Del 7-R |

Similarly, two-step PCR was used to make fragments for these constructs using plasmid pNoIP as a template. Del 1-F and Del 1-R were used to make the Del 1 construct; Del 1-F and Del 2-R were used to make the Del 2 construct; Del 1-F and Del 3-R were used to make the Del 3 construct; Del 4-F and Del 4-R were used to make the Del 4 mutant; Del 4-F and Del 5-R were used to make the Del 5 mutant; Del 6-F and Del 6-R were used to make the Del 6 mutant and Del 4-F and Del 7-R were used to make the Del 7 mutant.
Similarly, construct fragments were made by two-step PCR. All mutants were generated using pNOIP plasmid as the template, excepting DelG and the double mutant DelG Del α, which were generated using wild-type (pTPP-WT) plasmid as the template. The wild-type plasmid contained the full-length P3 stem and the first intron. To construct the M1 mutant, M1-5′F and M1-5′R were used to make the 5′ mutation first and then M1-3′F and M1-3′R were used to make the 3′ mutation to obtain the M1 mutant. The M4 mutant was constructed similarly. M2-F and M2-R were used to make the M2 mutant and M5-F and M5-R were used to make the M5 mutant. The M3 mutant contained both the M1 and M2 mutations and M6 contained both the M4 and M5 mutations. Del α F and Del α R were used to make the Del α mutant and the Del G Del α double mutant. Del G F and Del G R were used to make the mutant Del G.

| DNA29  | 5′-CAAGGTTAGATGAGAGGTGTCCTGCACTACGCGGCGCCGAG | M1-5′F |
|DNA30  | 5′-CTCGGGCCGGCGATGAGGCGAAGCTCTCTCATCTAACCCTT | M1-5′R |
|DNA31  | 5′-CTTTCGAGAGACGGGCCTGGCCGATCTATCTCCCC | M1-3′F |
|DNA32  | 5′-GGGGAATAGATCGGACCCAGAGCCTGGTTTCTACGCAAG | M1-3′R |
|DNA33  | 5′-GCAACATGAAGATCGAGCGCTAGCAACCGGACACGAG | M2-F |
|DNA34  | 5′-CTTGTCGTTGCTGCTAGGCGGCTCTGATCTTCTTGC | M2-R |
|DNA35  | 5′-GATGAGAGGTTGCCAGTGCTTACGCGGCGGCGAGTC | M4-5′F |
|DNA36  | 5′-GACTCGGCGCCGTAGACTGCGCAACCTCTCATC | M4-5′R |
|DNA37  | 5′-GCTTGCGAGAAACGAGCGAGCCGCGCTCCGAATCTTAT | M4-3′F |
|DNA38  | 5′-GAATAGATCGGACCGGCTCGTGTTTCTACGCAAGC | M4-3′R |
|DNA39  | 5′-GAAGATACAGAGCGGTTCGTACGACGCGAATCAGAG | M5-F |
|DNA40  | 5′-GGCTCTGGTTGCTGTACGGGACACGCGCTAGCTTTC | M5-R |
|DNA41  | 5′-ACGACCGGAACAGCAGCGATTTC | Del α F |
|DNA42  | 5′-GAAATGCTCTGTTCGCTGTTTCTACGCAAGC | Del α R |
|DNA43  | 5′-GTTGCGGCTCACTAGCGCGCGGCGAGTCAAACGCTTG | Del G F |
|DNA44  | 5′-CAAGCCTTTGGACTCGGCGCGAGTGAGCCGCAAC | Del G R |

| RT-PCR primers for WT |
|-----------------------|
| DNA45  | 5′-ATGGGAACCCCGTGTTGC | RT-F |
| DNA46  | 5′-ACAGGTATAGATGGTGATTACAC | RT-R |
| DNA47  | 5′-CATGGCCCTTGGAGCTGTC | Luci-RT-R |

Primers RT-F and RT-R were used for RT-PCRs in Fig. 1; Primers RT-F and luci-RT-R were used for Fig. 5.