A mechanism for ligand gated strand displacement in ZTP riboswitch transcription regulation

Eric J. Strobel[1]*, Luyi Cheng[2,4], Katherine E. Berman[2,4], Paul D. Carlson[3,4], Julius B. Lucks[1,4]*

Affiliations:
[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA
[2]Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL 60208, USA
[3]Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, 14853, USA
[4]Center for Synthetic Biology, Northwestern University, Evanston, IL, 20208

*Correspondence to: Email: eric.strobel@northwestern.edu, jblucks@northwestern.edu

Abstract

Cotranscriptional RNA folding forms structural intermediates that can be critically important for RNA biogenesis. This is especially true for transcriptional riboswitches that must undergo ligand-dependent structural changes during transcription to regulate the synthesis of downstream genes. Here, we systematically map the folding states traversed by the *Clostridium beijerinckii pfl* riboswitch as it controls transcription termination in response to the purine biosynthetic intermediate ZMP. We find that after rearrangement of a non-native hairpin to form the ZTP aptamer, cotranscriptional ZMP binding stabilizes two structural elements that lead to antitermination by tuning the efficiency of terminator hairpin nucleation and strand displacement. We also uncover biases within natural ZTP riboswitch sequences that could avoid misfolded intermediates that disrupt function. Our findings establish a mechanism for ZTP riboswitch control of transcription that has similarities to the mechanisms of diverse riboswitches and provide evidence of selective pressure at the level of cotranscriptional RNA folding pathways.

Introduction

The coupling of transcription and RNA folding is a ubiquitous property of RNA biogenesis (Pan & Sosnick, 2006). Nascent RNA folding is directed by the 5' to 3' polarity of transcription and the typically slower rate of the nucleotide addition cycle relative to base pair formation (Al-Hashimi & Walter, 2008; Lai, Proctor, & Meyer, 2013; J. Zhang & Landick, 2016). Consequently, cotranscriptional RNA folding inherently favors the formation of local structures that can pose energetic barriers to the formation of long-range interactions (Wong, Sosnick, & Pan, 2007). Thus, the cotranscriptional folding of complex RNA structures is expected to involve encounters with kinetic traps that can direct or prevent folding of biologically functional states (Chauvier et al., 2017; Drogalis & Batey, 2018; Heilman-Miller & Woodson, 2003; Pan, Artsimovitch, Fang, Landick, & Sosnick, 1999; Perdrizet, Artsimovitch, Furman, Sosnick, & Pan, 2012; Watters, Strobel, Yu, Lis, & Lucks, 2016; Wong et al., 2007; Yu et al., 2018).

The tendency of RNA molecules to enter kinetic traps as they fold cotranscriptionally is thought to be the basis for gene regulation by riboswitches (Garst & Batey, 2009). Riboswitches are RNA genetic switches that adopt alternate conformations to control gene expression in response to chemical ligands (Garst, Edwards, & Batey, 2011; Roth & Breaker, 2009; Serganov & Nudler, 2013). The identification of riboswitch ligands has consistently revealed diverse roles for RNA molecules in cellular physiology and illustrated the capacity of RNA for precise chemical recognition (McCown, Corbino, Stav, Sherlock, & Breaker, 2017; Nelson & Breaker, 2017). In addition, riboswitches have found diverse utility as antibiotic targets (Howe et al., 2015), diagnostic biosensors (Kellenberger, Wilson, Sales-Lee, & Hammond, 2013; Porter, Polaski, Morck, & Batey, 2017), and as imaging tools (Braselmann et al., 2018).

The general architecture of a riboswitch comprises a ligand-sensing aptamer domain and an 'expression platform' that directs a regulatory outcome based on whether ligand occupies the aptamer (Batey, 2012). In some cases, these domains share an overlap sequence that participates in mutually exclusive aptamer and expression platform structures that either block or allow downstream gene expression (Garst et al., 2011; Roth & Breaker, 2009; Serganov & Nudler, 2013). For riboswitches that regulate transcription, the coupling of transcription dynamics and RNA folding is particularly important

for determining which state folds, since folding and ligand recognition must occur within the limited time window of transcription (Wickiser, Cheah, Breaker, & Crothers, 2005; Wickiser, Winkler, Breaker, & Crothers, 2005). For these riboswitches, ligand binding to the aptamer is thought to kinetically trap the aptamer fold, effectively sequestering the overlap sequence and preventing the expression platform from forming (Helmling et al., 2018; Perdrizet et al., 2012; Watters, Strobel, et al., 2016; Zhao, Guffy, Williams, & Zhang, 2017). While crystal structures of diverse ligand-bound riboswitch aptamers have provided a detailed understanding of ligand-aptamer complexes (McCown et al., 2017), the detailed sequence of RNA folding pathways that underlie the structural basis of riboswitch regulatory control has only been described in a handful of cases (Chauvier et al., 2017; Perdrizet et al., 2012; Watters, Strobel, et al., 2016). As a result, uncovering general structure-function principles for how ligand stabilization of specific aptamer elements can cause kinetic traps that bifurcate cotranscriptional folding pathways to mediate control of gene expression remains a major goal.

Here we sought to directly address this knowledge gap by investigating the cotranscriptional folding mechanism of a recently discovered riboswitch that senses the purine biosynthetic intermediate 5-aminoimidazole-4-carboxamide riboside 5'-triphosphate (ZTP). Several years ago, the widely distributed *pfl* RNA motif was identified as a riboswitch aptamer for ZTP that regulates genes associated with purine biosynthesis and one-carbon metabolism (Kim, Nelson, & Breaker, 2015). The identification of the ZTP riboswitch uncovered a mechanism by which ZTP and its monophosphate derivative ZMP (Figure 1A) serve as bacterial 'alarmones' for deficiency in the cofactor 10-formyl-tetrahydrofolate (10f-THF): ZTP accumulates in bacteria upon 10f-THF deficiency to signal the expression of genes associated with 10f-THF biosynthesis (Bochner & Ames, 1982; Kim et al., 2015). The ZTP aptamer comprises a helix-junction-helix motif (P1-J1/2-P2) and a small hairpin (P3) that are separated by a variable length linker but interact through space to form the ZTP binding pocket (Jones & Ferre-D'Amare, 2015; Kim et al., 2015; Ren, Rajashankar, & Patel, 2015; Trausch, Marcano-Velazquez, Matyjasik, & Batey, 2015) (Figure 1B). In the *Clostridium beijerinckii* (Cbe) *pfl* ZTP riboswitch, the intrinsic terminator overlaps the aptamer such that the 5' half of the terminator stem can form internal base pairs to comprise the P3 hairpin, the loop of which interacts with the upstream J1/2 junction to form a pseudoknot that associates the two aptamer domains (Kim et al., 2015) (Figure 1B).

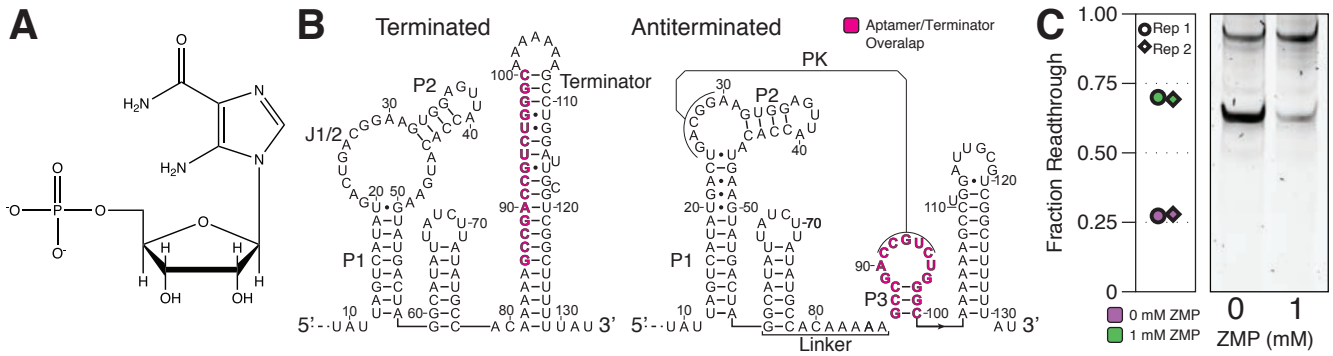


Figure 1. Overview of the *Clostridium beijerinckii pfl* ZTP riboswitch. (A) Chemical Structure of ZMP. **(B)** Secondary structure models of the terminated and antiterminated folds derived from covariation analysis (Kim et al., 2015) and x-ray crystallography (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015). Three paired regions (P1-P3), the terminator hairpin, the junction between P1 and P2 (J1/2), and the linker between the P1-J1/2-P2 and P3 aptamer subdomains are annotated. The aptamer/terminator overlap sequence is highlighted in magenta. **(C)** *In vitro* transcription in the presence and absence of 1 mM ZMP. Reaction conditions match those used for cotranscriptional SHAPE-Seq. Fraction read through values are quantified from the gel image. n=2 independent biological replicates are shown.
Figure 1-source data 1. Source data for Figure 1 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).

Formation of this pseudoknot then sequesters the terminator at both the secondary and tertiary levels (Figure 1B). Crystallographic studies of several ligand-bound ZTP aptamers revealed an extensive network of contacts between the aptamer subdomains that could stabilize the aptamer against transcription termination (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015). However the ZMP-dependence of these interactions during transcription remains unclear and the complexity of the aptamer sub-domain interface obscures the precise mechanism through which aptamer state dictates a cotranscriptional folding outcome. We therefore sought to understand how specific ligand-dependent contacts within the ZTP aptamer form cotranscriptionally to define its function as a transcription antiterminator.

In this work, we first characterize ligand-dependent folding pathways for the Cbe *pfl* ZTP riboswitch using high-throughput cotranscriptional RNA structure probing. Our analysis revealed two key features of *pfl* aptamer folding pathway: 1) the aptamer fold is preceded by a non-native intermediate that is potentiated by low-diversity aptamer sequences and 2) organization of the overall aptamer tertiary fold is ZMP-independent but ZMP binding stabilizes a network of discrete tertiary structure interactions that disfavor terminator nucleation and folding. The observation of ZMP-independent aptamer folding raises two important questions: how is the apo ZTP aptamer disrupted to allow terminator formation in the absence of ZMP, and how does the holo aptamer disfavor termination? To answer these questions, we devised a high-throughput mutagenesis approach to

identify the pathway by which transcription terminator nucleation unfolds the ZTP apo aptamer and identify key sequence determinants that govern this process. An analysis of diverse ZTP riboswitch sequences further revealed that a critical switching element of the ZTP aptamer has context-dependent sequence preferences that avoid deleterious off-pathway folded states that interfere with proper function. Overall our mechanistic findings of the ZTP riboswitch are remarkably similar to those found or proposed for disparate riboswitch systems, pointing to the generality of mechanisms of these important genetic switches.

Results

A non-native intermediate fold precedes nascent ZTP aptamer folding

We first sought to map the cotranscriptional folding pathway of the ZTP riboswitch in the presence and absence of ZMP to uncover key folding intermediates that facilitate transcription antitermination. To do this, we used cotranscriptional Selective 2'-Hydroxyl Acylation analyzed by Primer Extension Sequencing (SHAPE-Seq), which couples high-throughput chemical RNA structure probing (reviewed in (Strobel, Yu, & Lucks, 2018) and (Bevilacqua & Assmann, 2018)) with roadblocked *in vitro* transcription to comprehensively map possible intermediate nascent RNA folds in the context of transcription elongation complexes (TECs) (Strobel, Watters, Nedialkov, Artsimovitch, & Lucks, 2017; Watters, Strobel, et al., 2016). Cotranscriptional SHAPE-Seq begins with the stable distribution of TECs across every position of a DNA template using a multiplexed roadblocking strategy. After transcription initiation, nascent elongation complexes are treated with benzoyl cyanide (BzCN) (Mortimer & Weeks, 2007), which covalently modifies the 2'-hydroxyl of flexible RNA nucleotides, before the transcription reaction is terminated and chemical adducts are detected by reverse transcription and Illumina sequencing (Watters, Yu, Strobel, Settle, & Lucks, 2016).

To identify intermediate folds of the Cbe *pfl* ZTP riboswitch we performed cotranscriptional SHAPE-Seq with 0 mM or 1 mM ZMP using *E. coli* RNA polymerase (Figure 2). The *pfl* riboswitch was previously shown to be functional when *in vitro* transcribed by *Escherichia coli* RNA polymerase with limiting NTPs (Kim et al., 2015), and we were able to confirm riboswitch function in the saturating NTP concentration used for cotranscriptional SHAPE-Seq (Figure 1C).

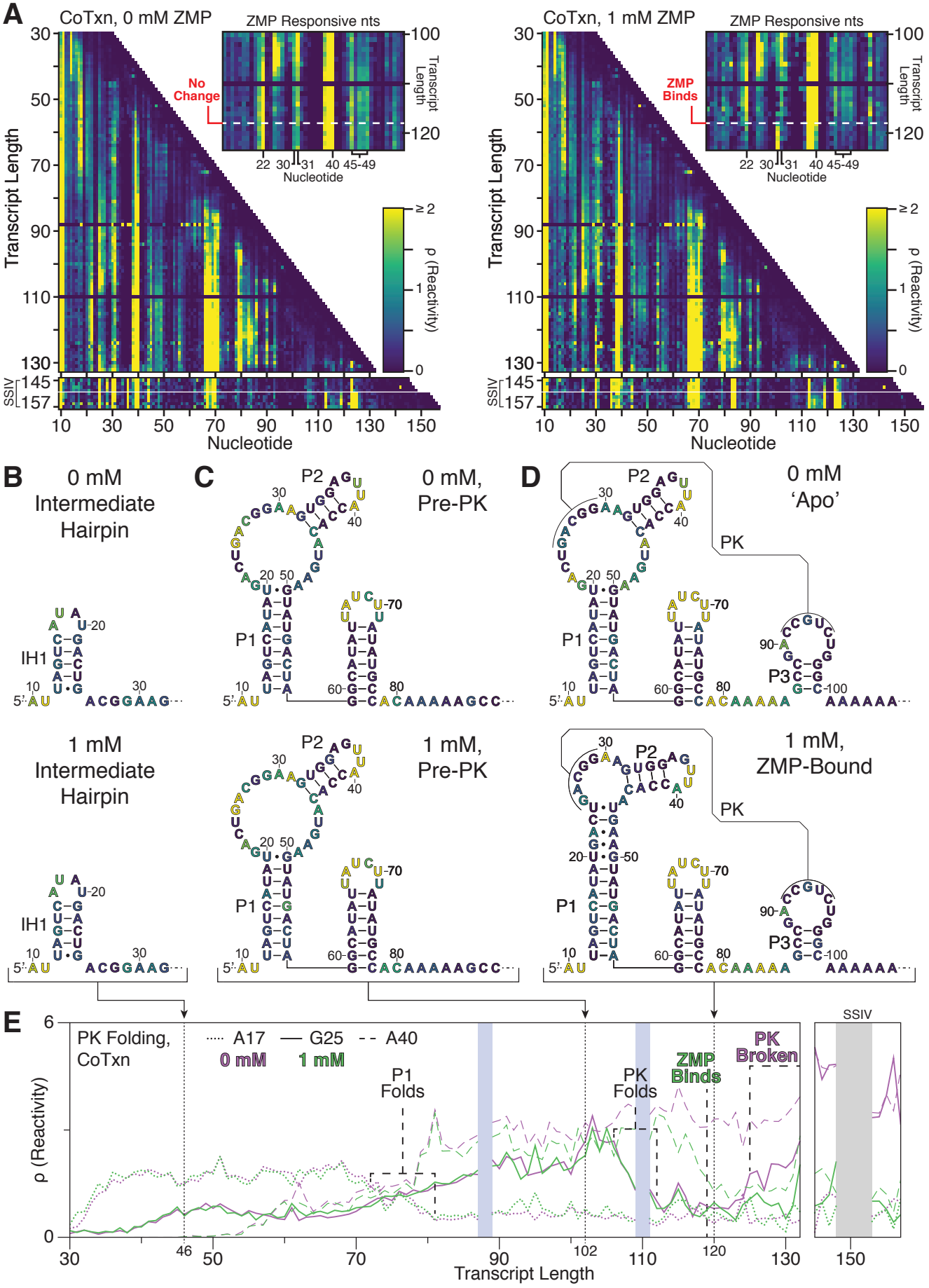

A
CoTxn, 0 mM ZMP
ZMP Responsive nts
No Change
Transcript Length
Nucleotide
ρ (Reactivity)
SSIV
CoTxn, 1 mM ZMP
ZMP Binds
B
0 mM
Intermediate
Hairpin
IH1
C
0 mM,
Pre-PK
P1
P2
D
0 mM
'Apo'
PK
P3
1 mM
Intermediate
Hairpin
1 mM,
Pre-PK
1 mM,
ZMP-Bound
E
PK Folding,
CoTxn
A17
G25
A40
0 mM
1 mM
P1
Folds
PK
Folds
ZMP
Binds
PK
Broken
Transcript Length

Figure 2. Chemical probing reveals ligand-independent and dependent intermediate folds along the *Clostridium beijerinckii pfl* ZTP riboswitch cotranscriptional folding pathway. (A) Cotranscriptional SHAPE-Seq reactivity matrix for the *Clostridium beijerinckii pfl* ZTP riboswitch with 0 mM and 1 mM ZMP. ZMP-responsive nts are highlighted in reactivity matrix cut-outs. The absence of data for transcripts 88 and 110 is due to ambiguous alignment of 3' end sequencing reads. Reactivity spectra for transcripts 145-148 and 153-157 are from a separate experiment using terminal biotin-streptavidin roadblocks and Superscript IV reverse transcriptase (RT) to resolve elevated RT stalling in transcripts beyond the terminator. **(B)** Intermediate hairpin (IH1) secondary structure identified by manual analysis of reactivity values and refined with minimum free energy structure predictions of the leader region. Nucleotides are colored by the reactivity of transcript length 46 from (A). **(C)** Secondary structure of P1-J1/2-P2 and the linker region based on the *pfl* RNA motif structure (Kim et al., 2015). Nucleotides are colored by the reactivity of transcript length 102 from (A). **(D)** Apo and ZMP-bound secondary structure as determined by manual SHAPE-refinement of the consensus structure (Apo) (Kim et al., 2015) or by the ZTP aptamer crystal structures (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015), respectively. Nucleotides are colored by the reactivity of transcript length 120 from (A). Nucleotides at the RNA 3' end of RNA structures that lie within the RNAP footprint are omitted in panels (B-D). **(E)** ZTP riboswitch folding as depicted by reactivity trajectories for nucleotides A17 (within IH1 loop or P1 stem, dotted lines), G25 (within J1/2 or the pseudoknot, solid lines), and A40 (within P2 loop, dashed lines) from (A). Trajectories are shown for both the absence (purple) and presence (green) of 1mM ZMP. Notable structural changes are identified by vertical dashed lines and the position of critical intermediate structures in (B-D) are shown by vertical dotted lines. Shaded boxes indicate the regions of poor alignment described in (A).
Results shown are from one representative of n=3 independent biological replicates (A, upper matrices; E) or n = 2 independent biological replicates (A, lower matrices); Replicate data are shown in Figure 2 supplement 5.
Figure 2-source data 1. Source data for Figure 2 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).
Figure 2 supplement 1. SHAPE probing of equilibrated *Cbe pfl* ZTP riboswitch intermediates.
Figure 2 supplement 2. Analysis of ZTP riboswitch sequences for intermediate hairpin formation.
Figure 2 supplement 3. Cotranscriptional SHAPE probing of disrupted pseudoknot mutant.
Figure 2 supplement 4. Cotranscriptional SHAPE probing of restored pseudoknot mutant.
Figure 2 supplement 5. Cotranscriptional SHAPE probing of the Cbe pfl ZTP riboswitch Replicate Data

Initial folding of the *pfl* aptamer is defined by the ZMP-independent formation of an intermediate hairpin (IH1) from nts 12-25 at transcript length 45 that comprises non-native base pairs between P1 and J1/2 (Figure 2B). IH1 persists through transcripts ~45 to ~70, during which P2 folds as indicated by the appearance of high reactivity at nts 38-40 (Figure 2A). At transcript length ~70 a gradual decrease in IH1 loop reactivity inversely correlates with increased reactivity throughout J1/2, suggesting that the emergence of P1 nucleotides 50-58 from RNAP drives refolding into the mutually exclusive P1 fold (Figure 2C, E). The mutual exclusivity of IH1 and P1 is further supported by SHAPE probing of equilibrium refolded *pfl* riboswitch intermediates, which reveals a sharp delineation between the IH1 and P1 structures across transcripts 55 and 56 and suggests that IH1 is a thermodynamically favorable folding intermediate until P1 can form. (Figure 2-figure supplement 1).

Because cotranscriptional SHAPE-Seq probes RNAs within roadblocked elongation complexes, it is possible that IH1 folding is enabled by transcription arrest. However, because the timescale of formation of local RNA structures is typically orders of magnitude faster than nucleotide addition by bacterial RNA polymerases (Mustoe, Brooks, & Al-Hashimi, 2014; Pan & Sosnick, 2006), the persistence of IH1 for at least 35 nt addition cycles suggests that IH1 folds even when transcription is uninterrupted. To investigate the potential prevalence of non-native structures in this early riboswitch

region, we assessed the capacity for hairpin formation in this sequence region of 532 ZTP riboswitch sequences from fully sequenced bacterial genomes (Kim et al., 2015). Given our recent observation that cotranscriptionally-generated folds of short RNA sequences match closely to those from equilibrium refolded experiments (Yu et al., 2018), we used RNAstructure's Fold method (Reuter & Mathews, 2010) to predict minimum free energy structures. We found that ~50% of predicted minimum free energy structures of this region within the natural sequence pool were more energetically favorable than *Cbe* IH1 (Figure 2-figure supplement 2A). Randomized sequences constrained by the nucleotide frequency observed in the ZTP riboswitches sampled above approximated the distribution of free energies observed in native ZTP riboswitch sequences, suggesting that the presence of intermediate structures in this region is a consequence of the natural nucleotide distribution (Figure 2-figure supplement 2B). Overall these analyses combined with probing data suggest that transient formation of a non-native intermediate hairpin prior to P1 folding is a feature of some ZTP riboswitches that is favored by the sequence constraints of the aptamer.

ZTP aptamer pseudoknot folding is independent of ZMP Binding

Following rearrangement of IH1 to form P1, we next observed the folding of a hairpin from nts 59-78 and an adjacent 7 nt unstructured region that comprise a linker between P1 and P3 (Figure 2C). ZMP-independent pseudoknot folding is then observed as a decrease in reactivities at nts 25-29 across transcript lengths 106-112, when nts 92-95 are expected to have emerged from RNAP given the ~14 nt footprint of RNAP on nascent RNA (Komissarova & Kashlev, 1998) (Figure 2E). Equilibrium refolding also revealed abrupt pseudoknot formation at transcript 95, when complete pairing between nts 25-29 and 91-95 first becomes possible (Figure 2-figure supplement 1), which is consistent with the coordinates of pseudoknot folding identified by cotranscriptional measurements since there is no RNAP footprint in equilibrium refolding experiments. Together, these data indicate that the pseudoknot can fold before the P3 stem forms, though ZMP-dependent reactivity changes are only observed after the entire P3 stem has emerged from RNAP around transcript length 117 to form the complete aptamer structure (Figure 2A, D, E). Consistent with this observation, SHAPE probing of equilibrium refolded RNAs reveals ZMP-dependent reactivity changes precisely when the P3 stem is expected to fold at

transcripts 99 and 100 (Figure 2-figure supplement 1). Furthermore, mutations that disrupt and restore the pseudoknot disrupt and restore observed pseudoknot folding and ZMP-binding, respectively (Figure 2-supplements 3 and 4). The requisite folding of P3 prior to the observation of any ZMP-dependent reactivity differences is consistent with the observation that the P3 stem stacks with the highly conserved G97 to form the floor of the ZMP binding pocket (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015). This is similar to the role that the P1 helix of purine riboswitches plays in ligand recognition (Batey, 2012; Batey, Gilbert, & Montange, 2004). Thus, organization of the overall ZTP aptamer tertiary fold precedes ligand binding. Notably, we and others also found that the *B. cereus crcB* fluoride riboswitch similarly adopts its overall tertiary fold prior to ligand recognition (Watters, Strobel, et al., 2016; Zhao et al., 2017), hinting at a potential generality of this feature of the mechanism of riboswitches that bind ligands near pseudoknot motifs (Jones & Ferre-D'Amare, 2017).

ZMP binding is associated with coordinated multi-point aptamer stabilization

The most notable observations from our cotranscriptional SHAPE-Seq measurements are drawn from a comparison of the ‘apo’ and ‘holo’ intermediates to identify ZMP-responsive nucleotides (Figure 2A and Figure 2-supplement 1A). The ZTP aptamer binds ZMP at the interface of the pseudoknot and P3, enabling a continuous helical stack by forming a Hoogsteen edge-Watson-Crick edge base pair with U96 (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015). The coordinated ZMP-dependent reactivity changes observed immediately following P3 stem folding are consistent with equilibrium measurements made by in-line probing (Kim et al., 2015) and the ZMP-bound crystal structures (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015) and can be categorized as P1-associated or pseudoknot-contacting (Figure 3).

In both cotranscriptional and equilibrium SHAPE experiments we observe a coordinated ZMP-dependent reactivity decrease across nts G21, A22, A45 and nts 47-49 at transcript lengths 117 and 100, respectively, indicating reduced flexibility within this region in the ZMP-bound aptamer (Figure 3A, Figure 3-figure supplement 1A). Disruption of pseudoknot pairing renders the P1-associated nucleotides ZMP-nonresponsive (Figure 3-figure supplement 1A). In the wild type riboswitch, ZMP-dependent reduced reactivity in these P1-associated nucleotides suggest that ZMP-binding enables

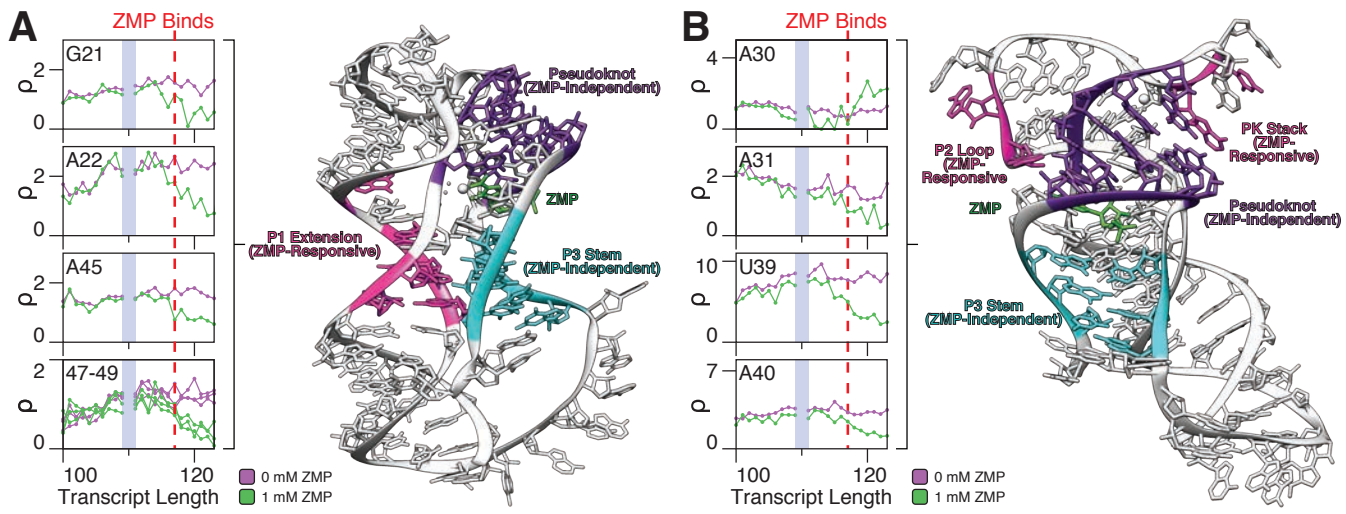


Figure 3. ZMP-responsive nucleotides reveal multipartite aptamer stabilization. ZMP-responsive nucleotides within the *Cbe* ZTP riboswitch, including **(A)** non-canonical P1 base pairs and **(B)** the nucleotides that interact with the pseudoknot, are shown in the absence (purple) and presence (green) of 1mM ZMP. Data are from figure 2A (wt *Cbe* cotranscriptional SHAPE-Seq replicate 1). The vertical red dashed lines indicate the transcript length at which ZMP binding is inferred to occur based on bifurcation of the reactivity trajectories. The crystal structure of the *Thermosinus carboxydivorans* ZMP riboswitch (PDB: 4ZNP) (Ren et al., 2015) is highlighted to show ZMP-stabilized P1 nuceotides (A) and ZMP-dependent pseudoknot-interacting nts (B). ZMP-responsive nts are highlighted in magenta, the pseudoknot is highlighted in purple, the P3 stem is highlighted in cyan, and ZMP is highlighted in green. Results shown are from one representative of n=3 independent biological replicates.
Figure 3-source data 1. Source data for Figure 3 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).
Figure 3 supplement 1. SHAPE reactivity of ZMP-responsive nucleotides in equilibrated wt and cotranscriptionally folded pseudoknot mutants.

stable base-pair formation between nucleotides 21-24 and 46-49, which form a primarily non-Watson-Crick (WC) helical extension of P1 (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015) (Figure 2D). This extension links P1 and P2 in a continuous stack and participates in a ribose-zipper interaction with P3 in the ZMP-bound crystal structures (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015) (Figure 3A).

The pseudoknot-contacting group comprises nts A31, U39, and A40, which show ZMP-dependent reduced reactivity in cotranscriptionally and equilibrium folded conditions at transcript lengths 117 and 100, respectively, (Figure 3B, Figure 3-figure supplement 1B). As above, disruption of the pseudoknot removes all ZMP-responsiveness in these nucleotides (Figure 3-figure supplement 1B). Decreased reactivity at U39 and the highly conserved A40 is likely attributable to the formation of a Type I A-minor interaction between A40 and the J1/2:P3 pseudoknot that was observed in the ZMP-bound crystal structures (Jones & Ferre-D'Amare, 2015; Kim et al., 2015; Ren et al., 2015; Trausch et al., 2015). A31 is not conserved in identity but has been shown to stack with the G26:C94 pseudoknot base pair in a ZMP-bound structure (Ren et al., 2015). The formation of this interaction could explain

the reduced reactivity at A31 and increased reactivity at A30 if stacking between A31 and the G29:C91 pseudoknot base pair places A30 in a bulge.

The ZMP-dependent stabilization of contacts between A40/A31 and each edge of the pseudoknot suggests that ZMP-binding coordinates several structural features of the aptamer: a P2 conformation that favors winding of the non-WC P1 base pairs, which in turn may favor the formation of the P1:P3 ribose zipper (Figure 3A). While these features are coordinated, our data cannot determine whether these happen in concert or in a series of folding events. Given that the entire *pfl* riboswitch P3 stem-loop participates in base pairs within the upstream portion of the terminator helix in the OFF state (Figure 1B), stabilization of P3 would be expected to disfavor terminator fold nucleation in the presence of ZMP, thus biasing the folding pathway into the ON state.

ZMP binding determines whether the antiterminated fold persists or is disrupted by the terminator

Transcribing RNAPs encounter the primary termination site at nt 132, only ~15 nts after signatures of ZMP binding are observed (Figure 2E). The first observed signature of terminator folding is a gradual increase in reactivity at the J1/2 pseudoknot nts 25-29 from transcript lengths ~125-132 in the absence of ZMP, suggesting that pseudoknot disruption can begin shortly after nucleotides within the 3' terminator stem have emerged from RNAP (Figure 2A,E). In contrast, sustained low reactivity at J1/2 in the presence of ZMP suggests that the pseudoknot remains stable (Figure 2A,E). The sole deviation from this trend was the primary termination site at transcript length 132 where transcripts that terminate naturally or as a consequence of roablocking at the termination site lead to increased J1/2 reactivity even in the presence of ZMP (Figure 2E). This is a technical limitation that we previously observed in our analysis of the fluoride riboswitch system, where a terminated fold was observed at the termination site even in the presence of fluoride (Strobel et al., 2017; Watters, Yu, et al., 2016).

With respect to terminator folding, we can observe signatures of fully terminated transcripts at length 132, which display a sharp reduction in reactivity across nts 81-93 (Figure 2A) indicating that the terminator hairpin is extended by bases pairs between the poly-U tract and the poly-A linker between P1 and P3 as RNA is extracted from RNAP (Figure 2A, Figure 1B). Under equilibrium refolding

conditions, terminator folding is observed as a stepwise process as terminator nucleotides are added to the transcript: from transcript length 113 to 118 displacement of pseudoknot base pairs increases reactivity at nts 25-29 one-by-one, which is observed as a diagonal line of increasing reactivity at these positions in the reactivity matrices (Figure 2- Supplement 1). In addition, from transcript lengths 121 to 129 reactivity across nts 90 back to 81 is reduced in a step-wise cross diagonal pattern as the remaining terminator stem folds and the poly-U pairs with the poly-A P1/P3 linker (Figure 2-Supplement 1). We note that the poly-A terminator hairpin loop remains unreactive in all conditions.

Observation of the final riboswitch structural states was obscured due to high levels of Superscript III reverse transcriptase stalling in transcripts beyond the terminator. To overcome this, we performed a separate experiment that enriched for transcript lengths 145 and 155 and used Superscript IV to improve full-length cDNA yield (Figure 2A). Consistent with the observation that different RTs have distinct adduct detection biases, we find that SSIV reactivity measurements are sparse but in qualitative agreement with measurements made using SSIII (Novoa, Beaudoin, Giraldez, Mattick, & M., 2017; Sexton, Wang, Rutenberg-Schoenberg, & Simon, 2017)(Figure 2A). The reduced reactivity at G25, A31, and A40 observed with SSIII in the presence of ZMP after transcript length 117 is recapitulated across all post-termination site lengths with SSIV, suggesting that the ZMP-bound ON state persists beyond the point at which the terminator can form when the RNA is cotranscriptionally folded (Figure 2A). In the absence of ZMP, high reactivity at G25, A31, A40 suggests that the apo aptamer pseudoknot is disrupted, leading to the terminated OFF state (Figure 2A, E). In contrast, equilibrium refolded conditions always show the disruption of the pseudoknot when it begins to compete with the terminator after transcript length 113, independent of ligand, indicating that the terminator fold outcompetes ligand-bound aptamer fold under refolding conditions (Figure 2-figure supplement 1). This is similar to our finding for the fluoride riboswitch (Watters, Strobel, et al., 2016) and suggests that the ZTP riboswitch antiterminated fold is only accessible by a cotranscriptional folding regime.

Optimal ZTP riboswitch sequence variants balance aptamer and terminator pairing states

Our analysis of ZTP riboswitch folding suggests a model in which (1) the pseudoknot folds independently of, but is required for, ZMP-binding, (2) ZMP-binding favors antitermination by stabilizing P3 through a series of interactions including the P1:P3 ribose zipper, and by stacking with ZMP (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015), and (3) termination occurs in the absence of ZMP through pseudoknot disruption. To validate these findings, we developed a high-throughput combinatorial mutagenesis strategy to specifically perturb RNA folding using multiplexed mutagenesis, transcription, and sequencing analysis of transcription products of targeted riboswitch variants (Figure 4-figure supplement 1A). Our strategy was inspired by a recently described comprehensive analysis of glycine riboswitch point mutations (Torgerson, Hiller, Stav, & Strobel, 2018) and other approaches that systematically perturb RNA transcripts (Vvedenskaya et al., 2015). To ask how specific groups of nucleotides interact during riboswitch folding we generated DNA templates from long synthetic oligonucleotides with randomization at several defined positions. A similar mutagenesis approach was recently described as part of an in-cell fluorescence-based genetic screen for riboswitch function (Polaski, Kletzien, Drogalis, & Batey, 2018). Using our mutant libraries, we performed a multiplexed *in vitro* transcription reaction in the presence and absence of ZMP, and RNA 3' end distributions were mapped by reverse transcription and high-throughput sequencing of transcription products to calculate the fraction of transcripts that read through past the terminator. Using this approach, we find that sequencing measurements approximate those made by gel electrophoresis and are highly reproducible (Figure 4-figure supplement 1B-E).

A key structural feature of many riboswitches is that the aptamer and expression platform structures are mutually exclusive (Figure 1B). This naturally creates an aptamer/terminator overlap sequence that can participate in two different structures depending on the overall riboswitch folding state. For the *pfl* ZTP riboswitch, this sequence comprises the entire P3 stem, which in the case of the aptamer fold participates in the pseudoknot interaction, and in the case of the terminator fold forms the 5' side of the terminator hairpin (Figure 1B).

We first sought to investigate how variation in this overlap sequence governs the competition for its folding context within the ZTP riboswitch cotranscriptional folding pathway. To do this, we constructed a mutagenesis library that varied Y91 and Y94 (Y=U,C) of the P3 loop and the respective

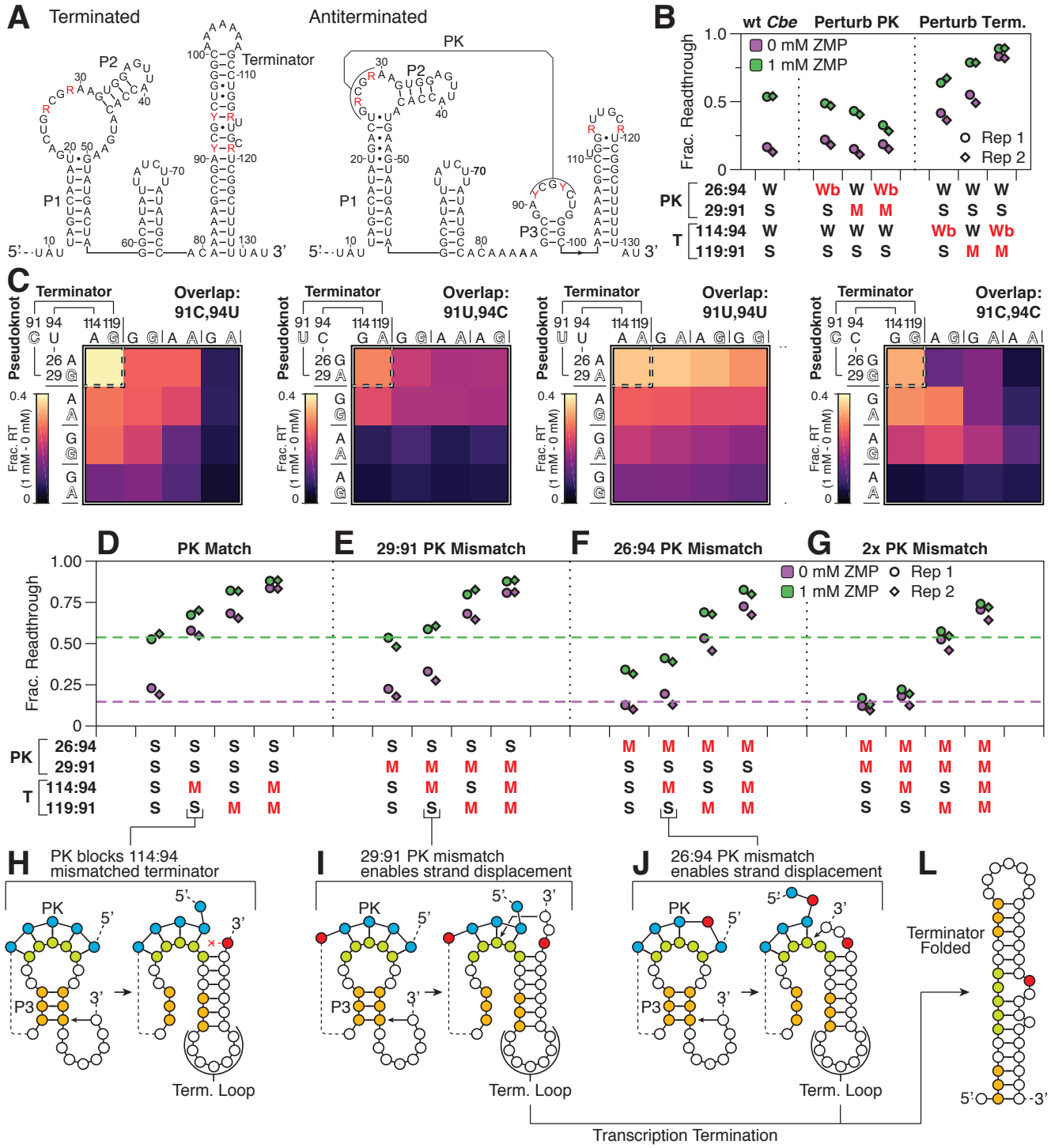


Figure 4. Combinatorial mutagenesis of the *Cbe pfl* ZTP riboswitch pseudoknot and terminator overlap region reveals sequence determinants of efficient terminator formation. (A) *Cbe* ZTP riboswitch terminated and antiterminated secondary structures depicting the randomization scheme used for aptamer/terminator overlap combinatorial mutagenesis. Randomized nucleotides are shown in red. R=A,G. Y=C,U. **(B)** Plot of fraction readthrough measured from high-throughput sequencing-based transcription assays for pseudoknot mutants as measured in the absence (purple) and presence (green) of 1 mM ZMP. Variant pairing patterns in the pseudoknot (PK) and terminator (T) are annotated as weak (W, A-U), strong (S, G-C), wobble (Wb, G-U), or mismatch (M, A-C). Red annotations indicate deviations from the wild-type pairing pattern. **(C)** Heat maps showing the difference in fraction readthrough between 1 mM and 0 mM ZMP conditions for pseudoknot mutants categorized by the identity of the aptamer/terminator overlap sequence (nts 91 and 94). For each group, the mutant containing perfect Watson-Crick base pairing in both the pseudoknot and terminator are shown by a dashed box (upper left). Competing pseudoknot and terminator base pairs are shown in matching type. **(D-G)** Fraction readthrough for pseudoknot match (D), 29:91 mismatch (E), 26:94 mismatch (F), and 29:91/26:94 mismatch variants (G) with terminator perturbations. Sequences are annotated as in panel (B). Dashed lines indicate observed fraction readthrough for the wt *Cbe* ZTP riboswitch with 0 mM

(purple) and 1 mM (green) ZMP. **(H-J)** Models for rescue of a 114:94 terminator mismatch (H) by a 29:91 (I) or 26:94 (J) pseudoknot mismatch. **(L)** Folded terminator secondary structure model. n=2 independent biological replicates are annotated as 'Rep 1' and 'Rep 2' in panels B and D-G. Heatmaps in panel C are the average of n=2 independent biological replicates. Individual replicate values are compared in Figure 4-figure supplement 1C.
Figure 4-source data 1. Source data for Figure 4 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).
Figure 4 Supplement 1. Targeted *in vitro* transcription and data analysis controls.
Figure 4 Supplement 2. Complete aptamer/terminator overlap combinatorial mutagenesis data.
Figure 4 Supplement 3. *in vitro* transcription of *Cbe* ZTP riboswitch terminator variants

base pairing partners R29 and R26 in the pseudoknot (R=A,G) and R119 and R114 in the terminator (Figure 4A). In this way we varied the pseudoknot sequence according to the consensus, while preserving all nucleotides that are >97% conserved (Kim et al., 2015). Measurements on this library were then made in the presence and absence of 1 mM ZMP under *in vitro* transcription conditions identical to cotranscriptional SHAPE-Seq. These mutants were found to have a range of function, characterized by a difference of the fraction readthrough values in the two conditions (1mM – 0mM ZTP) to be between 3.7% to 39% (wild-type = 39%) (Figure 4-figure supplement 2A).

As anticipated, mutations that perturb the *pfl* aptamer pseudoknot reduce ZMP-responsiveness and mutations that perturb the terminator base pairs increase terminator readthrough (Figure 4B). Furthermore, the best performing variants contained complete or near-complete Watson-Crick pairing in both the pseudoknot and terminator (Figure 4C). In contrast, variants with mismatches or wobble pairs typically reduced ZMP-dependent terminator readthrough either by reducing ZMP-responsiveness or increasing background terminator readthrough (Figure 4C). Among the variants that have Watson-Crick matches within the pseudoknot, the one with the most substantial departure from the wt measurements reversed the native A26:U94/G29:C91 base pairs to G26:C94/A29:U91 and displayed elevated terminator readthrough in the absence and presence of ZMP (Figure 4-figure supplement 2C-E). Lastly, terminator efficiency in the absence of ZMP was dependent on several factors including the position and severity of perturbations; weak base pairs closer to the poly-U tract produced severe termination defects (Larson, Greenleaf, Landick, & Block, 2008) and strong competing pseudoknot base pairs reduced terminator efficiency Figure 4-figure supplement 2B). Overall, these data highlight the importance of balanced pairing strength between the mutually exclusive aptamer and terminator folds in order to optimize the function of the *pfl* riboswitch.

pfl riboswitch terminator folding requires efficient strand displacement of the pseudoknot

The *pfl* riboswitch folding pathway as measured by cotranscriptional SHAPE-Seq shows that the aptamer pseudoknot can fold in the absence of ZMP and is subsequently disrupted by transcription of the terminator sequence. Despite the lack of a defined nucleation point, we reasoned that the most likely pathway for terminator folding and consequent pseudoknot disruption would begin with strand displacement through the 3' side of P3 by the 3' side of the terminator as it is transcribed. To validate this hypothesis, we focused on a variant that had five GC pseudoknot pairs and asked how function was altered by the presence of mismatches in either the pseudoknot, the terminator, or both.

We began by considering variants with perfect pseudoknot pairing, in which even a single terminator mismatch increased terminator readthrough at 0 mM ZMP from 23% to 58-84% (Figure 4D). In the context of the 114A:94C mismatch, which is distal to the poly-U tract and therefore not expected to be profoundly deleterious to terminator function, a linear model for terminator folding where nucleation begins with base-pair formation proximal to the terminator loop would permit only one pseudoknot base pair to be disrupted before the mismatch is encountered (Figure 4H). This would leave four contiguous base pairs intact and thereby disfavor propagation of terminator base pairs. In contrast, the 119A:91C mismatch is not positioned to interfere with pseudoknot strand displacement but may cause a general termination defect due to its proximity to the poly-U tract, which was observed in all cases in which this mutation was present and is consistent with observations in the *Clostridium tetani* glycine riboswitch (Torgerson et al., 2018) and for transcription terminators in general (Larson et al., 2008).

We next asked whether pseudoknot mismatches could rescue function of terminator mismatches. An 29A:91C pseudoknot mismatch partially recovered the termination defect of a 114A:94C terminator mismatch from 58% to 33% terminator readthrough (Figure 4E), confirming that this mismatch does not cause an inherent terminator defect. The 29A:91C mismatch reduces the pseudoknot to four contiguous base pairs such that terminator folding would then reduce the pseudoknot to three base pairs before encountering a mismatch (Figure 4I, L). This three base pair pseudoknot could permit strand displacement, leading to a functional, but leakier switch than a perfectly paired terminator.

Similarly, a 26A:94C pseudoknot mismatch, which is proximal to the ZMP binding pocket, reduced ZMP responsiveness but recovered the 114A:94C mismatch terminator defect from 58% to 19% terminator readthrough in the absence of ZMP (Figure 4F). The 26A:94C pseudoknot mismatch not only reduces the pseudoknot to three contiguous base pairs but could also permit unimpeded propagation of terminator base pairs once the first pseudoknot base pair is broken by the strand displacement process (Figure 4J, L).

Finally, a doubly mismatched pseudoknot showed no ZMP responsiveness, but behaved identically to a 26A:94C mismatch in the absence of ZMP (Figure 4G). Interestingly, the 119A:91C terminator mismatch resulted in a broken ON variant in all cases tested, indicating that a mismatch in this portion of the terminator is deleterious to its function, consistent with its close proximity to the poly-U tract (Larson et al., 2008; Torgerson et al., 2018) (Figure 4D-G).

Taken together, these data support a model in which the apo pseudoknotted structure is broken by a terminator strand displacement process that is most efficient with Watson-Crick base pairing through its progression. Notably, the native *pfl* riboswitch includes several bulges in the terminator stem that could interfere with displacement of pseudoknot base pairs during terminator folding (Fig 1B). Removal of these bulges does not meaningfully impact termination efficiency in the absence of ZMP but reduces terminator readthrough in the presence of ZMP by ~15 percentage points, suggesting that imperfections during strand displacement can tune function (Figure 4-figure supplement 3).

Labile P3 base pairing is a critical determinant of termination efficiency in the absence of ZMP

Cotranscriptional SHAPE-Seq measurements identified the ZMP-dependent stabilization of P1 base pairs that directly contact P3 through the formation of a ribose zipper, implicating ZMP-mediated P3 stabilization as an important determinant of antitermination. To assess the role of P3 stem stability on termination efficiency and ZMP-dependent antitermination, we sought to design a random mutagenesis library to perturb the P3 stem. Because the P3 stem is highly conserved (Kim et al., 2015), we designed a library that extends P3 by two potential base pairs while preserving base pairing within the intrinsic terminator and remaining consistent with the ZTP aptamer consensus sequence (Figure 5A). The resulting 512 variants were designed to assess a specific model for terminator folding

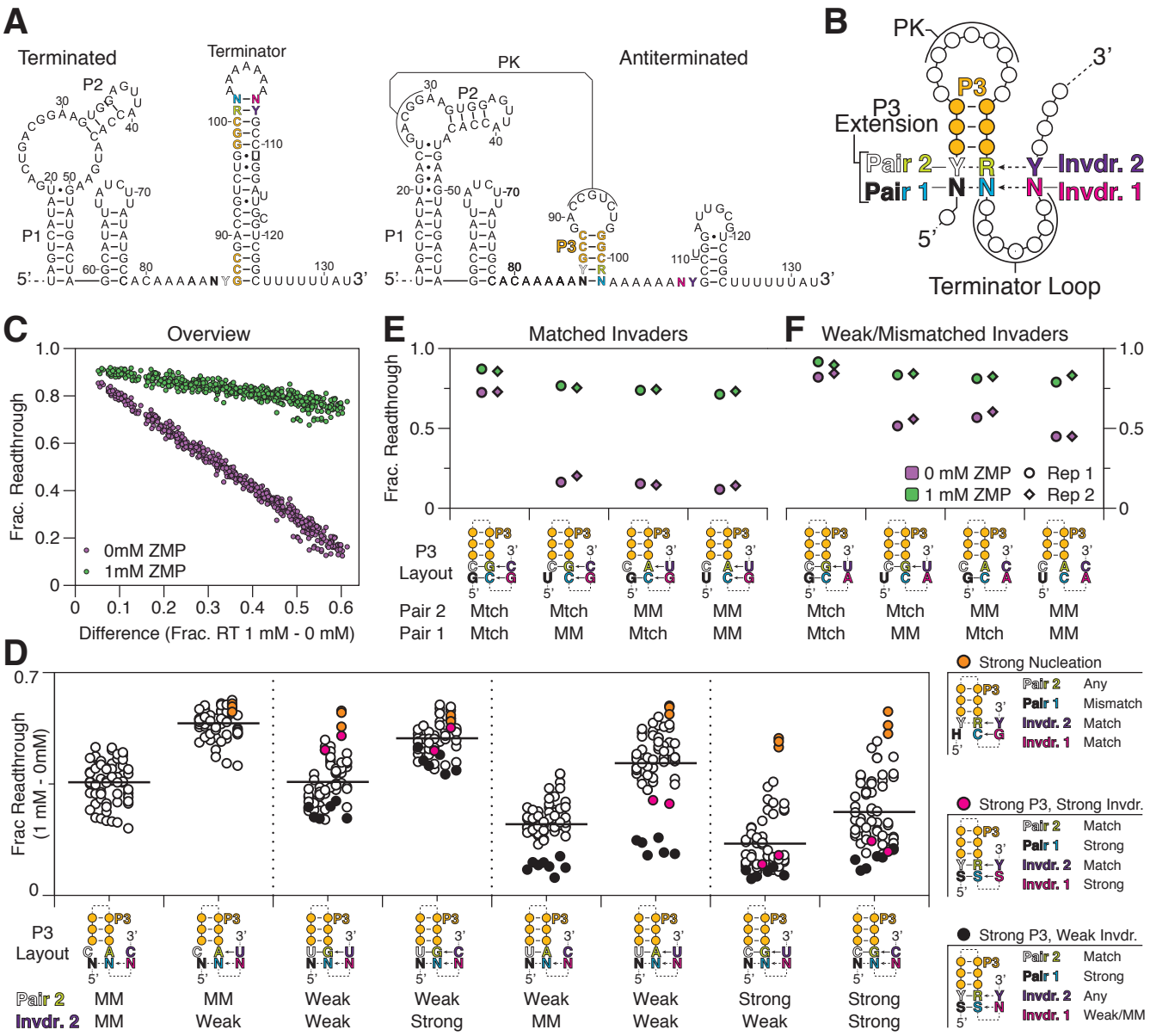


Figure 5. The P3 stem is a critical determinant of ZTP riboswitch termination efficiency.
(A) *Cbe* ZTP riboswitch terminated and antiterminated secondary structures depicting the randomization scheme used for P3 stem combinatorial mutagenesis. Mutations are color coded according to region. Insertion mutations are not counted in nucleotide numbering to maintain consistent nucleotide numbers. **(B)** Schematic of P3 stem/terminator base pairing competition. Randomized nucleotides are shown using the color codes in (A). Potential base pairs are shown as dashed lines or arrows. Invading nucleotides from the terminator are labeled Invdr. **(C)** Plot of fraction readthrough for P3 mutants as measured in the absence (purple) and presence (green) of 1 mM ZMP ordered by the difference in fraction readthrough observed in 1 mM and 0 mM ZMP conditions. **(D)** Plot of fraction readthrough (1 mM – 0mM) separated by Pair 2 and Invader 2 identity. Solid horizontal lines are the average fraction readthrough (1 mM – 0 mM) of the respective group. P3 Layout for each group is colored as in (B). Notable Pair 2/Invader 2 configurations are plotted in orange (strong nucleation), magenta (strong P3, strong invader), or black (strong P3, weak invader); variable P3 nucleotides are annotated as purine (R, A/G), pyrimidine (Y, U/C), strong (S, G/C), not G (H, A/U/C), or any (N, A/U/G/C). For these configurations, required base pairs are shown with a solid line and optional base pairs are shown in a dashed line. **(E-F)** Plot showing fraction readthrough for select P3 base pairing configurations with matched or weak/mismatched invading nucleotides. P3 Layout for each variant is shown as in (D). n=2 independent biological replicates are annotated as 'Rep 1' and 'Rep 2' in panels E and F. Panel C and D are the average of n=2 independent biological replicates. Individual replicate values are compared in Figure 4-figure supplement 1D.
Figure 5 source data 1. Source data for Figure 5 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).
Figure 5 Supplement 1. in vitro transcription of Cbe ZTP riboswitch toehold variants

that occurs by a process of nucleation and strand displacement which compete with existing base pairs at the base of the P3 stem (Figure 5B). In this model, nucleation and strand displacement must proceed by invading the bottom-most base pairs of P3 (Pair 1 and Pair 2) by two invading nucleotides (Invader 1 and Invader 2), which form the top of the terminator stem. In this way, termination efficiency is expected to be governed by P3 sequence and pairing composition.

The distribution of all variants ranked by the observed fraction read through in the presence and absence of ZMP indicated that all showed similar levels of terminator readthrough with ZMP, but a range of termination (10-80% fraction read through) in the absence of ZMP (Figure 5C). Thus, for this mutant group, ZMP-dependent activation is defined by termination efficiency. A closer examination of these values revealed several important trends (Figure 5D). First, classification of P3 variants by the strength of Pair 2, which can directly extend P3 by up to 1 bp, reveals that improved base pairing in P3 tends to reduce terminator efficiency (Figure 5D). Furthermore, for each Pair 2 variant, the presence of a complementary Invader 2 nucleotide improves termination efficiency in the absence of ZMP, and thus riboswitch function, relative to a mismatched Invader 2 (Figure 5D). Interestingly, we also observed that sequences in which a G in Invader 1 can make a strong pair with a 3' C side of Pair 1 always results in an optimally functioning variant within sequence groups, but not when the G-C pair is flipped in orientation (Figure 5D orange points,5E, Figure 5-figure supplement 1). A closer analysis of the interplay between several P3 base pairing configurations and perfectly matched invading nucleotides yield highly functional riboswitches except when P3 contains strong G-C pairs at its base that make it inaccessible to strand displacement (Figure 5E). Weak or mismatched invading pairs reduce termination efficiency in the absence of ZMP but remain sensitive to strong P3 base pairing in the same way (Figure 5F).

Overall, this data uncovers sequence features that promote or prevent efficient strand displacement into the terminated OFF state in the absence of ZMP and emphasize the accessibility of P3 to strand displacement as a critical determinant of termination efficiency. Combined with our analysis of pseudoknot pairing, these observations highlight the delicate balance between pairing interactions in the aptamer and the expression platform that allow ligand binding to tip the balance

towards the aptamer state in the cotranscriptional folding context, but still allow disruption of this state into the terminated fold when ligand is not present.

Context-dependent nucleotide frequency in natural ZTP riboswitch P3 sequences suggests a selective pressure for on-pathway folding

The analysis above supports a central role for P3 in ZTP riboswitch function as a core aptamer structure that is labile in the apo aptamer but becomes a barrier to terminator folding in the holo aptamer. Our functional analysis of P3 variants agrees with a previous analysis of ZTP riboswitch sequence composition which showed that 14 of 15 nucleotides that form P3/L3 are highly conserved in identity (Kim et al., 2015). The remaining nucleotide, which corresponds to nucleotide A90 of the *Cbe pfl* aptamer, was present in >97% of ZTP aptamer sequences but showed no conservation for identity. Structural studies showed this nucleotide in three distinct conformations, further supporting its apparent lack of conservation (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015). Despite this, given the importance of the P3 region for ZTP riboswitch function and because the *Thermosinus carboxydivorans* crystal structure showed the equivalent A nucleotide contacting the pseudoknot through an A:G Watson-Crick edge-Hoogsteen edge base pair (Figure 6-figure supplement 1A), we sought to investigate if changing the identity of this single non-conserved nucleotide would impact riboswitch function. We randomized nucleotide A90 of the *Cbe pfl* ZTP aptamer alongside the ZMP-responsive nucleotides A31 and A40 as part of a third mutagenesis experiment intended to perturb nucleotides that directly contact the pseudoknot (Figure 6, Figure 6-figure supplement 1D-G). To allow N90 complementation in the terminator hairpin we randomized U120, with the expectation that cognate nt 90:120 pairs would produce a functional terminator (Figure 6A). Counter to this expectation, all matched 90:120 pairs that differed from the wt A90:U120 pair increased terminator readthrough in the absence of ZMP compared to the wt, indicating that a match between these positions does not guarantee proper terminator function (Figure 6B). Indeed, 90Y:120R combinations functioned as poorly as 90:120 mismatches (Figure 6B, Figure 6-figure supplement 1C). These observations indicated that changing the identity of position 90 can have a large impact on *pfl* riboswitch function.

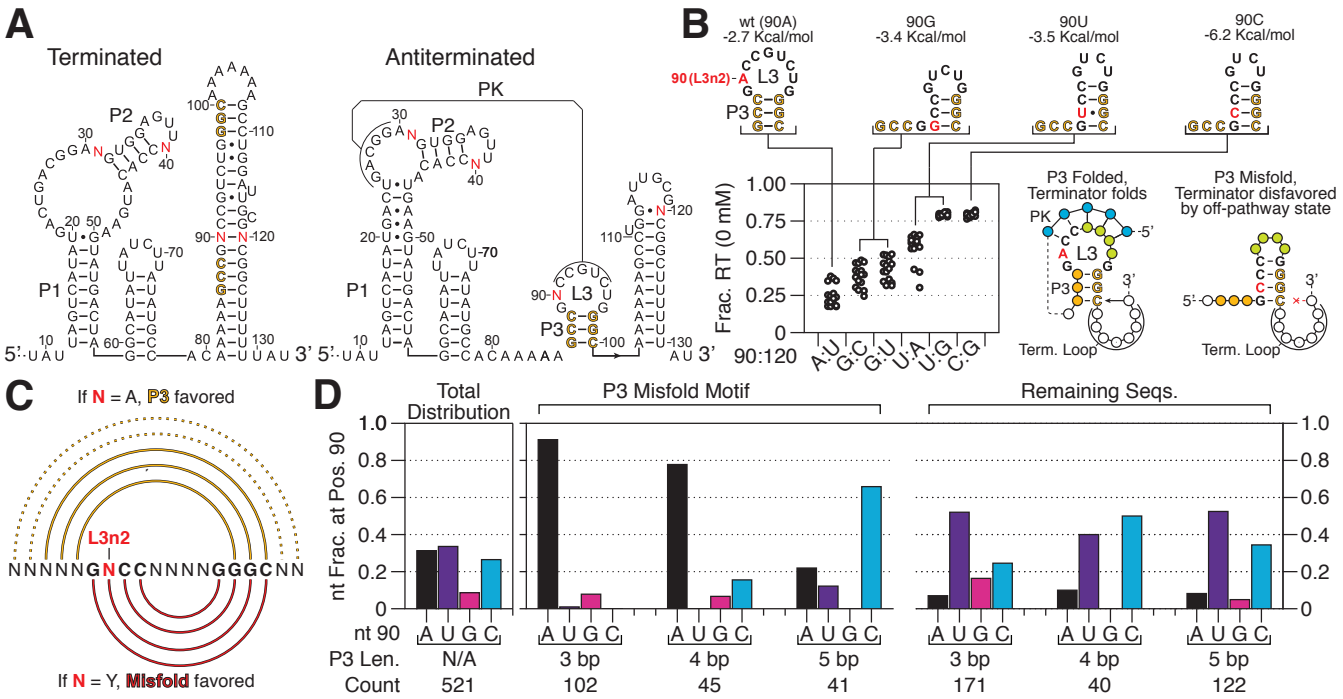


Figure 6. Context-dependent nucleotide frequency in P3 sequences suggests selective pressures for on-pathway folding.
(A) *Cbe* ZTP riboswitch terminated and antiterminated secondary structures depicting the randomization scheme used for combinatorial mutagenesis of nucleotides that contact the ZTP aptamer psuedoknot. P3 stem nucleotides are gold and randomized nucleotides are red. **(B)** Fraction readthrough in the absence of ZMP for all variants from (A) that can form a base pair between nucleotides 90 and 120. Structure predictions for the P3 region (nts 86-100) are shown for each N90 variant with an associated ΔG predicted by the RNAstructure Fold command. A model for how a P3 misfold could interfere with terminator folding is shown. **(C)** P3 folding in the context of the P3 misfold motif sequence. The sequence region comprising the misfold motif is shown in bold text with a defined sequence except for the red 'N', which designates L3 nucleotide 2 (L3n2) and corresponds to A90 in the wt *Cbe pfl* ZTP aptamer. A complete match to defined nucleotides designates a sequence as having the capacity to misfold if L3n2=Y. All nucleotides that were not specified when classifying P3/L3 sequences are labeled 'N'. The native P3 fold is shown above the sequence text in gold, where arcs between two bases indicate pairing: solid lines designate critical P3 base pairs and dashed lines designate optional P3 base pairs observed in some ZTP riboswitch sequences (Kim et al., 2015). The red lines below the sequence text illustrate the modeled misfold structure that is favored over the native P3 structure when N=Y. **(D)** Nucleotide frequency at position 90 for native ZTP riboswitch sequences in aggregate and binned by the presence of the 'misfold motif' defined in (C) and the length of the P3 stem. Sequences containing non-conserved insertions or a maximum contiguous P3 stem of less than three base pairs (11 of 532 sequences) were not considered. Panel B is the average of n=2 independent biological replicates. Individual replicate values are compared in Figure 4-figure supplement 1E.
Figure 6-source data 1. Source data for Figure 6 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).
Figure 6 supplement 1. Complete pseudoknot contact combinatorial mutagenesis data and targeted in vitro transcription analysis of pseudoknot contact variants.

Given our observation that P3 lability is critical to termination efficiency, we sought to understand how perturbations at nucleotides 90 and 120 could influence aptamer and terminator folding. Because our cotranscriptional SHAPE-Seq data suggests that P1 and the P1-P3 linker fold before P3 has emerged from RNAP, we reasoned that P3 folding is a predominantly local process. We therefore began by calculating minimum free energy structures for the P3/L3-forming region of the ZTP aptamer (nts 86-100). Under this simplified model, the nt 90 G, U, and C variants can form off-pathway hairpin structures that are more energetically favorable than the native P3 fold (Figure 6B).

Furthermore, the trend in their predicted free energy stabilities correlates with impairment of termination efficiency and each of these modeled folds are in direct competition with P3 (Figure 6B) and contains terminal G:C pairs that, based on our analysis of P3 above (Figure 5), could interfere with terminator nucleation. According to this simple model then, changing the identity of nucleotide 90 promotes an off-pathway fold that inhibits terminator formation.

We next asked whether the surrounding nucleotide context could govern a conservation pattern at nucleotide 90 that avoids off-pathway folding. In our simplified folding model, eight nucleotides that comprise part of L3 and the 3' side of P3 form a four base-pair off-pathway fold if the second L3 nucleotide (L3n2) is a pyrimidine. We therefore denoted this sequence as the 'misfold motif' (Figure 6C) and sought to investigate whether the misfold motif is avoided in natural ZTP riboswitch sequences. Starting with 532 ZTP riboswitch sequences from fully sequenced bacterial genomes (Kim et al., 2015), we binned each sequence by i) the presence of the seven nucleotides that yield the misfold motif if L3n2 is a pyrimidine (Figure 6C), and ii) the number of contiguous P3 pairs as a proxy for the favorability of the correct P3 fold, and then determined the nucleotide distribution at L3n2. In the aggregate set of ZTP riboswitch sequences A, U, and C were present at L3n2 at approximately equal frequencies. However, >91 % (93/102) sequences with a 3 bp P3 stem that would complete the misfold motif if L3n2=Y, contain an A at L3n2 and 99% contain an R (Figure 6D). The frequency of A at L3n2 was slightly reduced to >79% A for sequences with a 4 bp P3 stem that complete the misfold motif if L3n2=Y. Interestingly, sequences with a 5 bp P3 stem do not avoid the misfold motif and favor L3n2 pyrimidines comparably to sequences without any capacity for the misfold motif (>78% Y vs. >81% Y, respectively) (Figure 6D). Thus, the sequence requirements that avoid misfolding with a less favorable P3 stem may not be necessary if P3 is highly favorable. Taken together, these observations suggest that in specific sequence and structure contexts there is a selective pressure within L3 to prevent misfolding of a crucial riboswitch structure. This effect is washed out when analyzing ZTP riboswitch sequences in aggregate because different L3 sequences and P3 stem lengths break the observed conservation pattern.

Discussion

Here we have shown how intermediate folds of the *Cbe pfl* ZTP riboswitch mediate its function as a ZTP-dependent transcriptional switch. Nascent RNA structure probing revealed a positional roadmap of ZTP riboswitch folds comprising the formation of a non-native hairpin early in the folding pathway, followed by aptamer organization, ligand-mediated aptamer stabilization, and bifurcation into terminated or antiterminated folds (Figure 7). We found that a pseudoknot central to the ZTP aptamer tertiary fold is organized independently of ZMP, but that ZMP binding stabilizes a network of secondary and tertiary contacts. This observation raises two questions central to understanding riboswitch mechanisms: how does terminator folding overcome the barrier presented by the ZTP aptamer in the absence of ZMP?, and how does ZMP binding prevent the terminator from forming? Answers to both of these questions point to determinants within the critical overlap sequence that govern the progression or prevention of folding pathways that transit between the mutually exclusive aptamer or expression platform structures that determine the riboswitch regulatory outcome.

Our analysis of the Cbe *pfl* ZTP riboswitch revealed that terminator folding proceeds by strand displacement through the ZTP aptamer. Terminator folding is efficient if the pairing interactions in the overlap sequence, which comprises the pseudoknotted P3 loop and P3 stem, are balanced with terminator base pairing. Nucleotide compositions that favor either the aptamer or the terminator structure tip this balance and lead to non-functional switches (Figures 4,5). Interestingly, we observed that an otherwise stable P3 stem can be efficiently disrupted if terminator folding can be nucleated by the formation of a base pair at the 3' base of the stem (Figure 5). This is reminiscent of 'toehold' strand displacement in synthetic DNA hybridization systems where efficient strand displacement among DNA molecules is seeded through interactions between initially unpaired DNA regions (Seelig, Soloveichik, Zhang, & Winfree, 2006; D. Y. Zhang, Turberfield, Yurke, & Winfree, 2007). While this mode for nucleating strand displacement is local relative to classical 'toeholds', the underlying principle was also recently reported for the *pbuE* adenine riboswitch (Drogalis & Batey, 2018) and appears to be general. Overall these considerations between balance in pairing interactions between the two mutually exclusive riboswitch folds, and motifs that allow folding pathways to traverse barriers caused by imbalances, could constrain riboswitch sequence evolution and greatly inform riboswitch design.

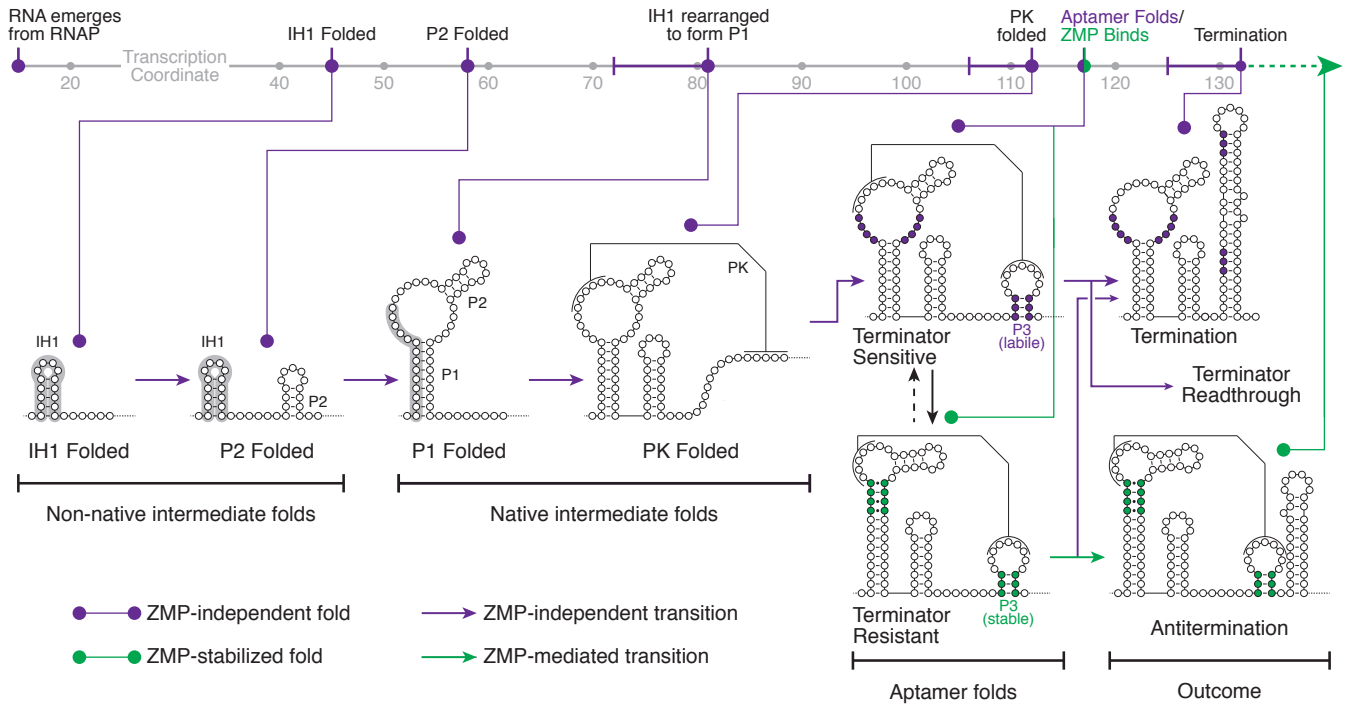


Figure 7. A model for ZTP riboswitch folding
ZTP riboswitch folding intermediates as observed by cotranscriptional SHAPE-Seq are plotted by transcription coordinate. Purple indicates ZMP-independent folds and transitions; Green indicates ZMP-stabilized folds or ZMP-mediated transitions. Initial ZTP aptamer folding comprises the formation of a non-native intermediate hairpin (IH1) that rearranges to form P1. Upon folding of both the pseudoknot (PK) and P3, the ZTP aptamer is competent to bind ZMP. In the absence of ZMP, terminator nucleation sequentially disrupts P3 and PK to fold the terminator hairpin and terminate transcription. In the presence of ZMP, the aptamer can become trapped in a stable fold that renders the P3 stem resistant to terminator nucleation, thereby driving antitermination.

Given that the *pfl* riboswitch terminator has evolved to efficiently strand displace through a pseudoknotted hairpin, how does ZMP binding favor antitermination? In agreement with equilibrium biochemical measurements (Kim et al., 2015), our analysis of nascent RNA structure shows that several ZMP-responsive nucleotides localize to a non-Watson-Crick extension of P1 that was found to form a ribose zipper with P3 (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015) (Figure 3). These interactions, along with the formation of a continuously stacked P3 stem-pseudoknot when ZMP pairs with U96 (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015), could disfavor terminator folding by stabilizing P3 against strand displacement. In the native *pfl* riboswitch, terminator nucleation competes directly with P3 base pairing and is therefore expected to be particularly sensitive to reduced P3 flexibility caused by these interactions. Thus, ZMP binding facilitates a versatile stabilization of P3 that can block both block terminator nucleation and strand displacement. This mechanism is remarkably similar to a model proposed for adenine-mediated stabilization of the *pbuE* aptamer P1 stem against terminator strand displacement (Batey, 2012).

An important constraint to ZTP-mediated antitermination is the narrow window of transcription coordinates in which the aptamer can bind ZMP before terminator folding. The *pfl* aptamer is competent to bind ZMP ~15 nts before RNAP reaches the primary termination site, but only ~7 nucleotides before terminator nucleation can begin to compete with the aptamer fold (Figure 2). ZMP binding within this 15 nt window could be promoted by several mechanisms. First, the ability of long-range pseudoknot base pairs to form up to 11 nucleotide addition cycles before P3 has fully folded provides time for tertiary organization of the aptamer before ligand recognition. Second, transcription pausing has been implicated in riboswitch folding and ligand recognition, and an appropriately positioned pause could extend the time for ZMP binding (Chauvier et al., 2017; Perdrizet et al., 2012). However, rigorous evaluation of pausing would require use of the cognate RNAP (Artsimovitch, Svetlov, Anthony, Burgess, & Landick, 2000). Finally, the observation that antitermination is efficient given sufficient ZMP suggests that ZTP aptamer folding is frequently successful, thus inefficiency in ligand binding may be tolerated over many attempted rounds of transcription.

Interestingly, our analysis of the *Cbe pfl* riboswitch also provides experimental evidence for two features that have been proposed to characterize cotranscriptional RNA folding pathways: the presence of temporary helices and the avoidance of nonfunctional competitor helices (Meyer & Miklos, 2004). Due to the rapid timescale of RNA folding with respect to transcription speed (Pan & Sosnick, 2006), local temporary helices may be expected to form when possible. However, local off-pathway folds are unexpected for riboswitch folding pathways since function requires successful aptamer folding and therefore mandates that non-native intermediates be resolved efficiently. Despite this, we observed that *pfl* riboswitch folding involves formation of the temporary helix IH1, which precedes P1 folding, making this one of the few examples of non-native intermediate folds have been experimentally observed (Fukuda et al., 2018; Incarnato et al., 2017; Watters, Yu, et al., 2016; Wong et al., 2007; Yu et al., 2018) (Figure 2). Interestingly, the IH1 structure is not explicitly encoded in the *pfl* aptamer consensus sequence. Rather, the tendency to form non-native structures is enriched by the low sequence diversity and high GC content of J1/2 and the separation of the sequences that comprise P1 by ~30 bp. These sequence features increase the probability that an intermediate fold can form and provide a long window for a structure to form, respectively. The simplicity of these sequence characteristics suggests

that IH1-like structures are likely to be prevalent in other non-coding RNAs. In addition to its capacity for intermediate temporary folds, the ZTP aptamer also exhibits a context-dependent sequence preference that avoids a misfolded alternative structure to P3 that would preclude ZMP recognition (Figure 6). Importantly, sequence preference at this position is dependent on both the potential for misfolding given the local sequence context and the favorability of the native P3 structure. While these findings are limited in scope, they support a general principle that RNA sequences are selected not just for folding into a final functional fold, but for folding into intermediate structures across their entire cotranscriptional folding pathways that efficiently guide them into those functional states (Lai et al., 2013).

Importantly, our mechanistic findings only apply to a cotranscriptional folding regime; equilibrium refolding favors terminator formation even in the presence of ZMP. Thus, the order of folding is essential for function because antitermination depends on cotranscriptional entry into a kinetically trapped holo aptamer fold. In this regard, the ZTP riboswitch behaves similarly to the *Bacillus cereus crcB* fluoride (Watters, Strobel, et al., 2016) and *Escherichia coli thiC* thiamine pyrophosphate (TPP) riboswitches (Chauvier et al., 2017). Notably, the *pfl* aptamer defaults to a pseudoknotted fold that requires the addition of only a single A:U or G:C pair to kinetically trap the riboswitch in an antiterminated state even in the absence of ligand binding. This is consistent with an analysis of the *Clostridium tetani* glycine riboswitch type-1 singlet which concluded that meaningful ligand-dependent changes in gene expression require finely tuned expression platform helices that are close in energy (Torgerson et al., 2018). Interestingly, in the case of the *pfl* aptamer, sequences that promote highly efficient terminator nucleation can circumvent an artificially stabilized aptamer. Thus, the *pfl* riboswitch folding pathway inherently assumes an ON decision and relies on both the thermodynamic favorability of the terminator structure, and the kinetic efficiency of strand displacement, to reject this structural assumption in the absence of ligand.

Perhaps most significant is the similarities and dissimilarities of mechanistic features of *pfl* riboswitch folding to other riboswitch systems. We previously observed that the cotranscriptionally folded *crcB* fluoride aptamer adopts a flexible psuedoknotted fold that is stabilized in the presence of fluoride (Watters, Strobel, et al., 2016). Thus, similar to the *pfl* ZTP riboswitch, the *crcB* fluoride riboswitch must efficiently exit this psuedoknotted fold without ligand but maintain it with ligand. Our

studies support a general mechanism to escape the aptamer fold through efficient strand displacement processes that are only blocked by multi-partite stabilizing interactions that form around the core pseudoknot when ligand is bound. The same mechanism has been proposed for the function of the pbuE adenine riboswitch (Batey, 2012; Drogalis & Batey, 2018) which uses a three-way junction aptamer fold to recognize its ligand. Notably, the apparent stability of the *pfl* aptamer pseudoknot in the absence of ZMP is distinct from the rapid docking and undocking dynamics observed between the *Faecalibacterium prausnitzii* class III preQ$_1$ riboswitch and its associated ribosome binding site (Liberman et al., 2015). Thus, whereas structurally distinct aptamers that recognize diverse ligands can use common strategies to make cotranscriptional genetic decisions, distinct folding dynamics appear to enable the function of diverse riboswitch platforms.

Materials and Methods

Standard DNA template preparation

Linear DNA template for *in vitro* transcription were prepared by PCR amplification as previously described (Strobel, Watters, Loughrey, & Lucks, 2016). Briefly, five 100 μl reactions containing 82.25 μl of water, 10 μl Thermo Pol Buffer (New England Biolabs, Ipswich, MA), 1.25 μl of 10 mM dNTPs (New England Biolabs), 2.5 μl of 10 μM oligonucleotide A (forward primer; Supplementary Table 4), 2.5 μl of 10 μM oligonucleotide B (reverse primer; Supplementary Table 4), 1 μl of Vent Exo- DNA polymerase (New England Biolabs), and 0.5 μl of plasmid DNA (Supplementary Table 3) were subjected to 30 cycles of PCR. Terminal biotin roadblock DNA templates were prepared as above but using either oligonucleotide D or oligonucleotide E (Supplementary Table 4) as the reverse primer. Reactions were pooled and precipitated by adding 50 μl of 3M sodium acetate (NaOAc) pH 5.5 and 1 mL of cold 100% ethanol (EtOH) and incubating at -80C for 15 min. After centrifugation, the precipitated pellet was washed once with 1.5 mL 70% EtOH (v/v), dried using a SpeedVac, and dissolved in 30 μl of water. The template was run on a 1% agarose gel and extracted using a QIAquick Gel Extraction kit (Qiagen, Hilden, Germany). DNA template concentration was determined by a Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA).

Biotinylated DNA template preparation

Linear DNA templates containing approximately 1 internal biotin modification were prepared by PCR amplification and gel extraction as previously described (Strobel et al., 2017). Briefly, PCR amplification was performed as described above except that in place of a standard dNTP mixture, each dNTP and corresponding biotin-11-dNTP was added to a total of 100 nmol such that ~1 biotin-11-dNTP is incorporated with in the transcribed region of each DNA duplex as described previously (Strobel et al., 2017). Biotin-11-dATP and biotin-11-dGTP were purchased from Perkin Elmer (Waltham, MA). Biotin-11-dCTP and biotin-11-dUTP were purchased from Biotium (Fremont, CA).

Combinatorial mutagenesis DNA template preparation

Linear DNA templates for combinatorial mutagenesis experiments were prepared by PCR amplification of gel-purified 'Ultramer' oligonucleotides (Integrated DNA technologies, Coralville, IA) (Supplementary Table 5). Ultramer oligonucleotides were first converted to double-stranded DNA in a 100 μl PCR containing 84.5 μl water, 10 μl 10X ThermoPol Buffer, 2 μl 10 mM dNTPs, 0.25 μl of 100 μM oligo F (forward primer; Supplementary Table 4), 0.25 μl of 100 μM oligo E (reverse primer; Supplementary Table 4), and 2 μl of Vent Exo- DNA polymerase. Thermal cycling was performed for 8 cycles. The primers in this reaction appended additional sequence to the ultramer. The PCR was purified using a QIAquick PCR purification kit (Qiagen) and quantified using a Qubit 3.0 Fluorometer. A second preparatory amplification was then performed using oligonucleotides G and E (Supplementary Table 4) followed by precipitation and gel extraction as described above for Standard DNA Templates. The resulting product was quantified using a Qubit 3.0 Fluorometer.

***in vitro* transcription (without roadblocking)**

Open promoter complexes were formed as previously described (Strobel et al., 2017) by incubating 100 nM DNA template and 2 U of *E. coli* RNAP holoenzyme (New England Biolabs) in transcription buffer (20 mM tris(hydroxymethyl)aminomethane hydrochloride (Tris-HCl) pH 8.0, 0.1 mM ethylenediaminetetraacetic acid (EDTA), 1 mM dithiothreitol (DTT) and 50 mM potassium chloride

(KCl)), 0.2 mg/ml bovine serum albumin (BSA), and 500 μM High Purity ATP, GTP, CTP and UTP (GE Life Sciences) for 10 min at 37C. When present, 5-aminoimidazole-4-carboxamide-1-β-D-ribofuranosyl 5'-monophosphate (ZMP; Sigma Aldrich, St. Louis, MO) at a stock concentration of 50 mM in dimethyl sulfoxide (DMSO) was added to a final concentration of 1 mM ZMP and 2% (v/v) DMSO. When ZMP was not included DMSO was added to 2% (v/v). Single-round transcription was initiated upon addition of $MgCl_2$ to 10 mM and rifampicin (Sigma Aldrich) to 10 μg/ml and a total reaction volume of 25 μl. Transcription proceeded for 30 s before 75 μl of TRIzol solution (Life Technologies) was added and RNAs were extracted according to the manufacturer's protocol.

***in vitro* transcription with random streptavidin roadblocks and chemical probing (cotranscriptional SHAPE-Seq)**

in vitro transcription for cotransctiptional SHAPE-Seq was performed as described (Strobel et al., 2017) for single-length DNA templates except that reactions were scaled to 50 μl and the open promoter complex protocol was modified to accommodate streptavidin binding as previously described (Strobel et al., 2017). Briefly, after incubation of reactions at 37C for 7.5 min to form open promoter complexes, SAv monomer (Promega, Fitchburg, WI) was then added to 2.5 μM and reactions were incubated for an additional 7.5 min before transcription was initiated by adding $MgCl_2$ to 10 mM and rifampicin (Sigma Aldrich) to 10 μg/ml for total reaction volume of 50 μl. Transcription proceeded for 30 s to permit stable distribution of transcription elongation complexes before reactions were split into 25 μl aliquots and mixed with 2.79 μl of 400 mM Benzoyl Cyanide (BzCN; Pfaltz & Bauer, Waterbury, CT) dissolved in anhydrous DMSO ((+) sample) or mixed with anhydrous DMSO ((-) sample) for ~2 s before transcription was stopped by the addition of 75 μl TRIzol solution (Life Technologies) and RNAs were extracted according to the manufacturer's protocol. DNA template was degraded by incubation in 20 μl 1x DNase I buffer (New England Biolabs) containing 1 U DNase I (New England Biolabs) at 37C for 30 min. 30 μl of water and 150 μl TRIzol were added to stop the reaction and RNAs were extracted according to the manufacturer's protocol. The resulting RNA was dissolved in 10 μl of 10% DMSO.

in vitro transcription for equilibrium refolding experiments were performed as above, except that all reactions contained ZMP to promote stable distribution of elongation complexes across all positions, transcription was stopped without performing chemical probing, and the resulting RNAs were resuspended in 25 μl water. Under this purification protocol the ZMP included during initial RNA synthesis should be completely depleted during the two subsequent phased extractions, as is evidenced by the difference between the equilibrium refolded and cotranscriptionally-folded matrices (compare Figure 2-figure supplement 1A). Equilibrium refolding was performed by denaturing at 95C for 2 min, snap cooling on ice for 1 min, and adding transcription buffer, 500 μM NTPs, 10 mM $MgCl_2$, and either ZMP to 1mM ZMP/2% DMSO or 2% DMSO before incubation at 37C for 20 min. SHAPE modification with BzCN was performed as described above before addition of 30 μl water and 150 μl TRIzol, extraction according to the manufacturer's protocol and resuspension in 10 μl of 10% DMSO.

in vitro transcription with a terminal roadblock (combinatorial mutagenesis)

We chose to include a terminal biotin-streptavidin roadblock as part of our combinatorial mutagenesis protocol so that terminator run-through RNAs would appear as cluster of 3' ends ~10 nts upstream of the DNA template end. After incubating reactions at 37C for 7.5 min to form open promoter complexes, SAv monomer (Promega) was then added to 2.5 μM and reactions were incubated for an additional 7.5 min before transcription was initiated by adding $MgCl_2$ to 10 mM and rifampicin (Sigma Aldrich) to 10 μg/ml for total reaction volume of 25 μl. Transcription proceeded for 30 s before 75 μl of TRIzol solution (Life Technologies) was added and RNAs were extracted according to the manufacturer's protocol, DNaseI treated, TRIzol extracted a second time as described above.

Gel Electrophoresis and Analysis of Gel Images

Extracted RNAs were fractionated by denaturing 7.0 M urea polyacrylamide gel electrophoresis (10 or 12% polyacrylamide). Gels were stained with SYBR Gold (Life Technologies), imaged with a Bio-Rad ChemiDoc Touch Imaging System, and quantified with Image Lab (Bio-Rad, Hercules, CA).

Fraction readthrough was determined by dividing the intensity of the terminator readthrough band by sum of the terminator readthrough and terminated bands.

Sequencing library preparation

Sequencing libraries for cotranscriptional SHAPE-seq were prepared either as previously described (Watters, Yu, et al., 2016) or with a modified protocol that uses Superscript IV (SSIV) for reverse transcription. All combinatorial mutagenesis libraries were prepared using the modified SSIV protocol. For convenience, all protocol modifications are described below in the context of the complete original protocol.

RNA 3' linker adenylation and ligation

5'-Phosphorylated linker (Oligonucleotide K, Supplementary Table 6) was adenylated using a 5' DNA Adenylation Kit (New England Biolabs) at 20x scale and purified by TRIzol extraction as previously described (Watters, Yu, et al., 2016). RNA 3' ligation was performed by first combining 10 μl extracted RNAs in 10% DMSO with 0.5 μl of SuperaseIN (Life Technologies), 6 μl 50% PEG 8000, 2 μl of 10X T4 RNA Ligase Buffer (New England Biolabs), 1 μl of 2 μM 5'-adenylated RNA linker, and 0.5 μl of T4 RNA ligase, truncated KQ (New England Biolabs) and mixing by pipetting. 0.5 μl of T4 RNA ligase 2, truncated KQ (New England Biolabs) was then added and the reaction was mixed by pipetting again and incubated at 25C for 3 hrs.

Reverse Transcription

Following linker ligation, RNAs were precipitated by adding 130 μl RNase-free water, 15 μl 3M NaOAc pH 5.5, 1 μl 20 mg/ml glycogen, and 450 μl of 100% EtOH and storing at -80C for 30 min. Following centrifugation, pellets were washed once with 500 μl 70% EtOH (v/v), residual ethanol was removed. For Superscript III reverse transcription, pellets were resuspended in 10 ul RNase-free water and 3 μl of reverse transcription primer (Oligonucleotide L, Supplementary Table 6) was added. Samples were then denatured at 95C for 2 min, incubated at 65C for 5 min, briefly centrifuged to pull all

liquid to the bottom of the tube, and placed on ice. 7 μl of Superscript III reverse transcription master mix (containing 4 μl of 5x First Strand Buffer (Life Technologies), 1 μl of 100 mM DTT, 1 μl 10 mM dNTPs, 0.5 μl RNase OUT (Invitrogen, Waltham, MA), and 0.5 μl Superscript III) was added to each sample and samples were mixed before being placed at 45C. Reactions were incubated at 45C for 1 min, 52C for 25 min, and 65C for 5 min to deactivate Superscript III. For Superscript IV reverse transcription, RNAs were precipitated as described for Superscript III but resuspended in 9.5 μl of RNase-free water before 3 μl of reverse transcription primer (Oligonucleotide L, Supplementary Table 6) was added. Samples were then denatured at 95C for 2 min, incubated at 65C for 5 min, briefly centrifuged to pull all liquid to the bottom of the tube, and placed on ice. 7.5 μl of Superscript IV reverse transcription master mix (containing 4 μl of 5x SSIV Buffer (Life Technologies), 1 μl of 100 mM DTT, 1 μl 10 mM dNTPs, 0.5 μl RNase OUT (Invitrogen), and 1 μl Superscript IV) was added to each sample and samples were mixed before being placed at 45C. Reactions were incubated at 45C for 1 min, 52C for 25 min, and 65C for 5 min, and 80C for 10 min to deactivate Superscript IV. Following heat inactivation of Superscript III or Superscript IV, 1 μl 4M sodium hydroxide (NaOH) was added and samples were heated at 95C for 5 min to hydrolyze RNA. Samples were then partially neutralized by the addition of 2 μl 1M hydrochloric acid (HCl) and precipitated by adding 69 μl 100% EtOH and storing at -80C for 15 min. Samples were centrifuged for 15 min at 4C and pellets were washed by adding 500 μl of 70% EtOH and inverting the tubes several times. After removing residual ethanol pellets were dissolved in 22.5 μl of RNase-free water.

Adapter ligation

Ligation of high-throughtput sequencing adapter was performed as previously described (Watters, Yu, et al., 2016). Briefly, the dissolved cDNA was mixed with 3 μl of 10x CircLigase Buffer (Epicentre), 1.5 μl of 50 mM $MnCl_2$, 1.5 μl of 1 mM ATP 0.5 μl of 100 μM DNA adapter (Oligonucleotide M, Supplementary Table 6), and 1 μl of CircLigase I (Epicentre, Madison, WI). Ligation reactions were incubated at 60C for 2 h and the ligase was heat inactivated by incubating at 80C for 10 min. DNA was precipitated by adding 70 μl nuclease-free water, 10 μl 3M NaOAc pH 5.5, 1 μl 20 mg/ml glycogen, and

300 μl of 100% EtOH and storing at -80C for 30 min before centrifugation. The pellets were dissolved in 20 μl of nuclease-free water, purified using 36 μl of Agencourt XP beads (Beckman Coulter, Brea, CA) according to the manufacturer's protocol, and eluted with 20 μl of 1X TE buffer.

Quality Analysis

Quality analysis for cotranscriptional SHAPE-Seq was performed as previously described (Watters, Yu, et al., 2016) by generating fluorescently labeled dsDNA libraries using oligonucleotides O, P, Q and R or S (Supplementary Table 6). Samples were analyzed by capillary electrophoresis using an ABI 3730xl DNA Analyzer. Combinatorial mutagenesis libraries were also assessed for quality in this way.

Sequencing Library Preparation

Sequencing library preparation for cotranscriptional SHAPE-Seq was performed as previously described (Watters, Yu, et al., 2016). Briefly, 3 μl of ssDNA library (+) and (-) channels was separately mixed with 33.5 μl of nuclease-free water, 10 μl 5x phusion Buffer (New England BioLabs), 0.5 μl of 10 mM dNTPs, 0.25 μl of 100 μM TruSeq indexing primer (Oligonucleotide T, Supplementary Table 6), 2 μl of 0.1 μM of channel-specific selection primer (Oligonucleotides R and S, Supplementary Table 6), and 0.5 μl of Phusion DNA polymerase (New England BioLabs). Amplification was performed with an annealing temperature of 65C and an extension time of 15 s. After 15 cycles, 0.25 μl of primer PE_F (Oligonucleotide Q, Supplementary Table 6) was added and libraries were amplified for an additional 10 cycles. Following amplification, libraries were allowed to cool to 4C completely before the addition of 0.25 μl of ExoI (New England Biolabs) and incubation at 37C to degrade excess oligonucleotides. ExoI was heat inactivated by incubating at 80C for 20 min. Libraries were then mixed with 90 μl of Agencourt XP beads (Beckman Coulter), purified according to the manufacturer's protocol and eluted in 20 μl of 1X TE buffer. The resulting sequencing libraries were quantified using a Qubit 3.0 Fluorometer (Life Technologies) and molarity was estimated using the length distribution observed in Quality Analysis. Sequencing libraries for combinatorial mutagenesis was performed as described above except that

(+)/(-) channel barcoding was arbitrarily assigned because each library was given a unique TruSeq barcode.

Cotranscriptional SHAPE-Seq Sequencing and analysis

Sequencing of cotranscriptional SHAPE-Seq libraries was performed by the NUSeq Core on an Illumina NextSeq500 using either 2x36 or 2x37 bp paired end reads with 30% PhiX. Cotranscriptional SHAPE-Seq data analysis was performed using Spats v1.0.1 as previously described (Watters, Strobel, et al., 2016) except that one mismatch was permitted during alignment following the observation that truncations at several nucleotides were enriched for a terminal mutation.

Combinatorial Mutagenesis Data Sequencing and Analysis

Sequencing of combinatorial mutagenesis libraries was performed on an Illumina MiSeq using a MiSeq Reagent Kit v3 (150-cycle). Libraries were loaded with a density of approximately 1000 K/mm^2 and sequenced with a cycle configuration of either Read1:37, Index:6, Read2:132 or Read1:35, Index:6, Read2:134 and included 10% PhiX. Asymmetrical reads were used so that nearly the entire riboswitch was included in read 2 and the transcript end was contained in read 1. Alignment was performed using custom software available at https://github.com/LucksLab/LucksLab_Publications/tree/master/Strobel_ZTP_Riboswitch. For mutants without insertions, reads were required to contain a perfect target match between nucleotides 11 and 130 of the riboswitch. Alignment to the riboswitch leader (nts 1-10) was not required because elevated Superscript IV dropoff in this A/U-rich region reduced the number of alignments, and omission of reads lacking this sequence recovered these reads but did not impact measurements (Figure 4- figure supplement 1F,G). Control analyses of non-insertion mutants that omitted the poly-uridine tract from alignment did not impact measurements (Figure 4- figure supplement 1F,G), therefore insertion mutants were only required to align through two poly-uridine tract nucleotides to permit usage of 150-cycle v3 MiSeq Reagent Kits. For all mutants, Single mismatches were permitted in Read 1 beyond nt 130 provided that the read could be unambiguously identified as terminated or full length. Reads that mapped between positions 130 and 134 were considered terminated and reads that mapped to position

>=135 were considered antiterminated. Fraction terminator readthrough was calculated for each variant by dividing the number of antiterminated reads by the sum of terminated and antiterminated reads. Terminator efficiency is equal to 1 – fraction terminator readthrough.

RNA structure prediction

RNA structure prediction was performed using the RNAstructure v6.1 (Reuter & Mathews, 2010) Fold command with default settings. For IH1 (Figure 2-figure supplement 2), wt sequences from previously identified ZTP riboswitches (Kim et al., 2015) were obtained from the RefSeq database (Pruitt, Tatusova, & Maglott, 2007) and aligned using INFERNAL v1.1.2 (Nawrocki & Eddy, 2013). The segment used for structure prediction contains sequence from the beginning of the multiple sequence alignment through the last unstructured nt within J1/2 (as determined by the RNAstructure Fold command (Reuter & Mathews, 2010)). Unbiased randomization allowed an equal probability for all nucleotides at each position. wt nucleotide distribution biased randomization was performed using the observed nucleotide frequency for each position as measured for the wt ZTP riboswitch multiple sequence alignment. All randomized data sets match the length distribution observed in the wt ZTP riboswitch set.

L3n2 Nucleotide Frequency Analysis

P3 hairpin sequences from the multiple-sequence alignment described above that did not contain insertions were first binned by a match to the motif NNNNNGNCCNNNNGGGCNN and then binned by the number of predicted contiguous P3 base pairs. The frequency of each nucleotide at the position corresponding to nucleotide 90 of the *pfl* aptamer was then determined.

Code Availability

Spats v1.0.1 can be accessed at https://github.com/LucksLab/spats/releases/. Scripts used in data processing are located at https://github.com/LucksLab/Cotrans_SHAPE-Seq_Tools/releases/ and https://github.com/LucksLab/LucksLab_Publications/tree/master/Strobel_ZTP_Riboswitch.

Data Availability

Raw sequencing data that support the findings of this study have been deposited in the Sequencing Read Archive (SRA) (http://www.ncbi.nlm.nih.gov/sra) with the BioProject accession code PRJNA510362. Individual BioSample accession codes are available in Supplementary Table 1. SHAPE-Seq Reactivity Spectra generated in this work have been deposited in the RNA Mapping Database (RMDB) (http://rmdb.stanford.edu/repository/) with the accession codes listed below Sample details are available in Supplementary Table 2. Source data for all figures are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H). All other data that support the findings of this paper are available from the corresponding authors upon request.

Accession Codes

Raw sequencing data that support the findings of this study have been deposited in the Sequencing Read Archive (http://www.ncbi.nlm.nih.gov/sra) with the BioProject accession code PRJNA510362. Individual BioSample accession codes are available in Supplementary Table 1. SHAPE-Seq Reactivity Spectra generated in this work have been deposited in the RNA Mapping Database (RMDB) (http://rmdb.stanford.edu/repository/) (Cordero, Lucks, & Das, 2012) with the accession codes ZTPRSW_BZCN_0001, ZTPRSW_BZCN_0002, ZTPRSW_BZCN_0003, ZTPRSW_BZCN_0004, ZTPRSW_BZCN_0005, ZTPRSW_BZCN_0006, ZTPRSW_BZCN_0007, ZTPRSW_BZCN_0008, ZTPRSW_BZCN_0009, ZTPRSW_BZCN_0010, ZTPRSW_BZCN_0011, ZTPRSW_BZCN_0012, ZTPRSW_BZCN_0013, ZTPRSW_BZCN_0014, ZTPRSW_BZCN_0015, ZTPRSW_BZCN_0016. Sample details are available in Supplementary Table 2.

Acknowledgments

We thank Rob Batey, Chad Torgerson, Scott Strobel, and Christopher Jones for thoughtful discussions; Jim Brink and Steve Hockema for review of combinatorial mutagenesis alignment software; Ronald Breaker and Keith Corbino for sharing a ZTP aptamer multiple sequence alignment; Kyle Watters for sharing a script to download RefSeq database entries. This work was supported by an Arnold O. Beckman Postdoctoral Fellowship (to E.J.S.), a New Innovator Award through the National Institute of General Medical Sciences of the National Institutes of Health (grant no. 1DP2GM110828 to J.B.L.),

Searle Funds at The Chicago Community Trust (to J.B.L.) and by Grant Number T32GM008382 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

Author Contributions

Eric J. Strobel, Conceptualization, Methodology, Investigation, Validation, Formal analysis, Data Curation, Software, Writing-original draft, Writing-review and editing, Supervision, Project Administration, Funding acquisition. Luyi Cheng, Investigation, Validation, Writing-review and editing. Katherine E. Berman, Investigation, Validation, Writing-review and editing. Paul D. Carlson, Investigation, Validation, Writing-review and editing. Julius B. Lucks, Methodology, Writing-original draft, Writing-reviewing and editing, Supervision, Resources, Project Administration, Funding acquisition.

Competing Interests

No competing interests declared

References

Al-Hashimi, H. M., & Walter, N. G. (2008). RNA dynamics: it is about time. *Curr Opin Struct Biol, 18*(3), 321-329. doi:10.1016/j.sbi.2008.04.004

Artsimovitch, I., Svetlov, V., Anthony, L., Burgess, R. R., & Landick, R. (2000). RNA polymerases from Bacillus subtilis and Escherichia coli differ in recognition of regulatory signals in vitro. *J Bacteriol, 182*(21), 6027-6035.

Batey, R. T. (2012). Structure and mechanism of purine-binding riboswitches. *Q Rev Biophys, 45*(3), 345-381. doi:10.1017/S0033583512000078

Batey, R. T., Gilbert, S. D., & Montange, R. K. (2004). Structure of a natural guanine-responsive riboswitch complexed with the metabolite hypoxanthine. *Nature, 432*(7015), 411-415. doi:10.1038/nature03037

Bevilacqua, P. C., & Assmann, S. M. (2018). Technique Development for Probing RNA Structure In Vivo and Genome-Wide. *Cold Spring Harb Perspect Biol, 10*(10). doi:10.1101/cshperspect.a032250

Bochner, B. R., & Ames, B. N. (1982). ZTP (5-amino 4-imidazole carboxamide riboside 5'-triphosphate): a proposed alarmone for 10-formyl-tetrahydrofolate deficiency. *Cell, 29*(3), 929-937.

Braselmann, E., Wierzba, A. J., Polaski, J. T., Chrominski, M., Holmes, Z. E., Hung, S. T., . . . Palmer, A. E. (2018). A multicolor riboswitch-based platform for imaging of RNA in live mammalian cells. *Nat Chem Biol, 14*(10), 964-971. doi:10.1038/s41589-018-0103-7

Chauvier, A., Picard-Jean, F., Berger-Dancause, J. C., Bastet, L., Naghdi, M. R., Dube, A., . . . Lafontaine, D. A. (2017). Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nat Commun, 8*, 13892. doi:10.1038/ncomms13892

Cordero, P., Lucks, J. B., & Das, R. (2012). An RNA Mapping DataBase for curating RNA structure mapping experiments. *Bioinformatics, 28*(22), 3006-3008. doi:10.1093/bioinformatics/bts554

Drogalis, L. K., & Batey, R. T. (2018). Requirements for efficient cotranscriptional regulatory switching in designed variants of the Bacillus subtilis pbuE adenine-responsive riboswitch. *bioRxiv*. doi:10.1101/372573

Fukuda, S., Yan, S., Komi, Y., Sun, M., Gabizon, R., & Bustamante, C. (2018). Stabilizing Obligatory Non-native Intermediates Along Co-transcriptional Folding Trajectories of SRP RNA Affects Cell Viability. *bioRxiv*. doi:10.1101/378745

Garst, A. D., & Batey, R. T. (2009). A switch in time: detailing the life of a riboswitch. *Biochim Biophys Acta, 1789*(9-10), 584-591. doi:10.1016/j.bbagrm.2009.06.004

Garst, A. D., Edwards, A. L., & Batey, R. T. (2011). Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol, 3*(6). doi:10.1101/cshperspect.a003533

Heilman-Miller, S. L., & Woodson, S. A. (2003). Effect of transcription on folding of the Tetrahymena ribozyme. *RNA, 9*(6), 722-733.

Helmling, C., Klotzner, D. P., Sochor, F., Mooney, R. A., Wacker, A., Landick, R., . . . Schwalbe, H. (2018). Life times of metastable states guide regulatory signaling in transcriptional riboswitches. *Nat Commun, 9*(1), 944. doi:10.1038/s41467-018-03375-w

Howe, J. A., Wang, H., Fischmann, T. O., Balibar, C. J., Xiao, L., Galgoci, A. M., . . . Roemer, T. (2015). Selective small-molecule inhibition of an RNA structural element. *Nature, 526*(7575), 672-677. doi:10.1038/nature15542

Incarnato, D., Morandi, E., Anselmi, F., Simon, L. M., Basile, G., & Oliviero, S. (2017). In vivo probing of nascent RNA structures reveals principles of cotranscriptional folding. *Nucleic Acids Res, 45*(16), 9716-9725. doi:10.1093/nar/gkx617

Jones, C. P., & Ferre-D'Amare, A. R. (2015). Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains. *Nat Struct Mol Biol, 22*(9), 679-685. doi:10.1038/nsmb.3073

Jones, C. P., & Ferre-D'Amare, A. R. (2017). Long-Range Interactions in Riboswitch Control of Gene Expression. *Annu Rev Biophys, 46*, 455-481. doi:10.1146/annurev-biophys-070816-034042

Kellenberger, C. A., Wilson, S. C., Sales-Lee, J., & Hammond, M. C. (2013). RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. *J Am Chem Soc, 135*(13), 4906-4909. doi:10.1021/ja311960g

Kim, P. B., Nelson, J. W., & Breaker, R. R. (2015). An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol Cell, 57*(2), 317-328. doi:10.1016/j.molcel.2015.01.001

Komissarova, N., & Kashlev, M. (1998). Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc Natl Acad Sci U S A, 95*(25), 14699-14704.

Lai, D., Proctor, J. R., & Meyer, I. M. (2013). On the importance of cotranscriptional RNA structure formation. *RNA, 19*(11), 1461-1473. doi:10.1261/rna.037390.112

Larson, M. H., Greenleaf, W. J., Landick, R., & Block, S. M. (2008). Applied force reveals mechanistic and energetic details of transcription termination. *Cell, 132*(6), 971-982. doi:10.1016/j.cell.2008.01.027

Liberman, J. A., Suddala, K. C., Aytenfisu, A., Chan, D., Belashov, I. A., Salim, M., . . . Wedekind, J. E. (2015). Structural analysis of a class III preQ1 riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. *Proc Natl Acad Sci U S A, 112*(27), E3485-3494. doi:10.1073/pnas.1503955112

McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E., & Breaker, R. R. (2017). Riboswitch diversity and distribution. *RNA, 23*(7), 995-1011. doi:10.1261/rna.061234.117

Meyer, I. M., & Miklos, I. (2004). Co-transcriptional folding is encoded within RNA genes. *BMC Mol Biol, 5*, 10. doi:10.1186/1471-2199-5-10

Mortimer, S. A., & Weeks, K. M. (2007). A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc, 129*(14), 4144-4145. doi:10.1021/ja0704028

Mustoe, A. M., Brooks, C. L., & Al-Hashimi, H. M. (2014). Hierarchy of RNA functional dynamics. *Annu Rev Biochem, 83*, 441-466. doi:10.1146/annurev-biochem-060713-035524

Nawrocki, E. P., & Eddy, S. R. (2013). Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics, 29*(22), 2933-2935. doi:10.1093/bioinformatics/btt509

Nelson, J. W., & Breaker, R. R. (2017). The lost language of the RNA World. *Sci Signal, 10*(483). doi:10.1126/scisignal.aam8812

Novoa, E. M., Beaudoin, J., Giraldez, A. J., Mattick, J. S., & M., K. (2017). Best practices for genome-wide RNA structure analysis: combination of mutational profiles and drop-off information. *bioRxiv*. doi:10.1101/176883

Pan, T., Artsimovitch, I., Fang, X. W., Landick, R., & Sosnick, T. R. (1999). Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc Natl Acad Sci U S A, 96*(17), 9545-9550.

Pan, T., & Sosnick, T. (2006). RNA folding during transcription. *Annu Rev Biophys Biomol Struct, 35*, 161-175. doi:10.1146/annurev.biophys.35.040405.102053

Perdrizet, G. A., 2nd, Artsimovitch, I., Furman, R., Sosnick, T. R., & Pan, T. (2012). Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc Natl Acad Sci U S A, 109*(9), 3323-3328. doi:10.1073/pnas.1113086109

Polaski, J. T., Kletzien, O. A., Drogalis, L. K., & Batey, R. T. (2018). A functional genetic screen reveals sequence preferences within a key tertiary interaction in cobalamin riboswitches required for ligand selectivity. *Nucleic Acids Res, 46*(17), 9094-9105. doi:10.1093/nar/gky539

Porter, E. B., Polaski, J. T., Morck, M. M., & Batey, R. T. (2017). Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat Chem Biol, 13*(3), 295-301. doi:10.1038/nchembio.2278

Pruitt, K. D., Tatusova, T., & Maglott, D. R. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res, 35*(Database issue), D61-65. doi:10.1093/nar/gkl842

Ren, A., Rajashankar, K. R., & Patel, D. J. (2015). Global RNA Fold and Molecular Recognition for a pfl Riboswitch Bound to ZMP, a Master Regulator of One-Carbon Metabolism. *Structure, 23*(8), 1375-1381. doi:10.1016/j.str.2015.05.016

Reuter, J. S., & Mathews, D. H. (2010). RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics, 11*, 129. doi:10.1186/1471-2105-11-129

Roth, A., & Breaker, R. R. (2009). The structural and functional diversity of metabolite-binding riboswitches. *Annu Rev Biochem, 78*, 305-334. doi:10.1146/annurev.biochem.78.070507.135656

Seelig, G., Soloveichik, D., Zhang, D. Y., & Winfree, E. (2006). Enzyme-free nucleic acid logic circuits. *Science, 314*(5805), 1585-1588. doi:10.1126/science.1132493

Serganov, A., & Nudler, E. (2013). A decade of riboswitches. *Cell, 152*(1-2), 17-24. doi:10.1016/j.cell.2012.12.024

Sexton, A. N., Wang, P. Y., Rutenberg-Schoenberg, M., & Simon, M. D. (2017). Interpreting Reverse Transcriptase Termination and Mutation Events for Greater Insight into the Chemical Probing of RNA. *Biochemistry, 56*(35), 4713-4721. doi:10.1021/acs.biochem.7b00323

Strobel, E. J., Watters, K. E., Loughrey, D., & Lucks, J. B. (2016). RNA systems biology: uniting functional discoveries and structural tools to understand global roles of RNAs. *Curr Opin Biotechnol, 39*, 182-191. doi:10.1016/j.copbio.2016.03.019

Strobel, E. J., Watters, K. E., Nedialkov, Y., Artsimovitch, I., & Lucks, J. B. (2017). Distributed biotin-streptavidin transcription roadblocks for mapping cotranscriptional RNA folding. *Nucleic Acids Res, 45*(12), e109. doi:10.1093/nar/gkx233

Strobel, E. J., Yu, A. M., & Lucks, J. B. (2018). High-throughput determination of RNA structures. *Nat Rev Genet, 19*(10), 615-634. doi:10.1038/s41576-018-0034-x

Torgerson, C. D., Hiller, D. A., Stav, S., & Strobel, S. A. (2018). Gene regulation by a glycine riboswitch singlet uses a finely tuned energetic landscape for helical switching. *RNA, 24*(12), 1813-1827. doi:10.1261/rna.067884.118

Trausch, J. J., Marcano-Velazquez, J. G., Matyjasik, M. M., & Batey, R. T. (2015). Metal Ion-Mediated Nucleobase Recognition by the ZTP Riboswitch. *Chem Biol, 22*(7), 829-837. doi:10.1016/j.chembiol.2015.06.007

Vvedenskaya, I. O., Zhang, Y., Goldman, S. R., Valenti, A., Visone, V., Taylor, D. M., . . . Nickels, B. E. (2015). Massively Systematic Transcript End Readout, "MASTER": Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. *Mol Cell, 60*(6), 953-965. doi:10.1016/j.molcel.2015.10.029

Watters, K. E., Strobel, E. J., Yu, A. M., Lis, J. T., & Lucks, J. B. (2016). Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat Struct Mol Biol, 23*(12), 1124-1131. doi:10.1038/nsmb.3316

Watters, K. E., Yu, A. M., Strobel, E. J., Settle, A. H., & Lucks, J. B. (2016). Characterizing RNA structures in vitro and in vivo with selective 2'-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods, 103*, 34-48. doi:10.1016/j.ymeth.2016.04.002

Wickiser, J. K., Cheah, M. T., Breaker, R. R., & Crothers, D. M. (2005). The kinetics of ligand binding by an adenine-sensing riboswitch. *Biochemistry, 44*(40), 13404-13414. doi:10.1021/bi051008u

Wickiser, J. K., Winkler, W. C., Breaker, R. R., & Crothers, D. M. (2005). The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell, 18*(1), 49-60. doi:10.1016/j.molcel.2005.02.032

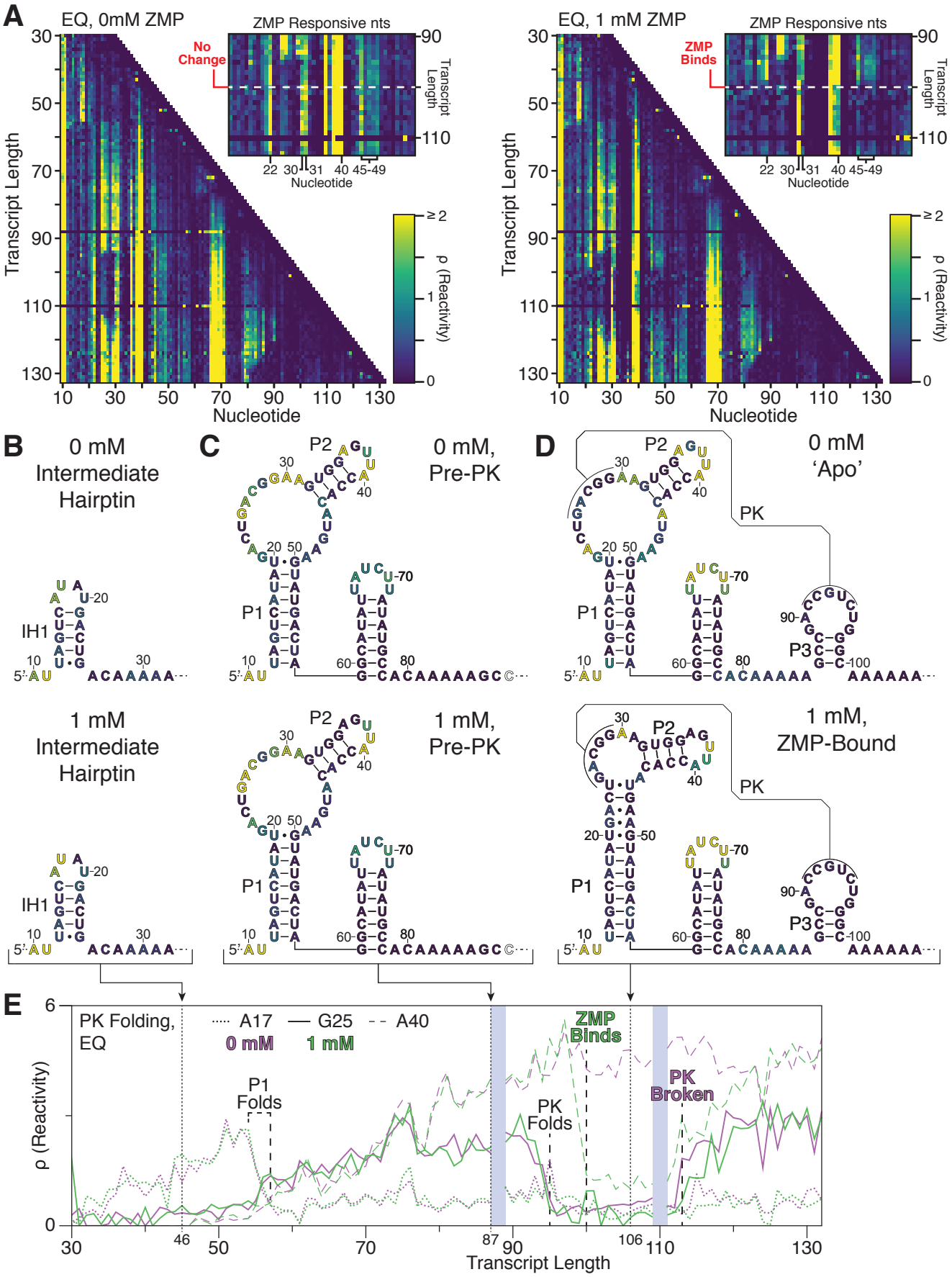
Wong, T. N., Sosnick, T. R., & Pan, T. (2007). Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc Natl Acad Sci U S A, 104*(46), 17995-18000. doi:10.1073/pnas.0705038104

Yu, A. M., Gasper, P. M., Strobel, E. J., Watters, K. E., Chen, A. A., & Lucks, J. B. (2018). Computationally Reconstructing Cotranscriptional RNA Folding Pathways from Experimental Data Reveals Rearrangement of Non-Native Folding Intermediates. *bioRxiv*. doi:10.1101/379222

Zhang, D. Y., Turberfield, A. J., Yurke, B., & Winfree, E. (2007). Engineering entropy-driven reactions and networks catalyzed by DNA. *Science, 318*(5853), 1121-1125. doi:10.1126/science.1148532

Zhang, J., & Landick, R. (2016). A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem Sci, 41*(4), 293-310. doi:10.1016/j.tibs.2015.12.009

Zhao, B., Guffy, S. L., Williams, B., & Zhang, Q. (2017). An excited state underlies gene regulation of a transcriptional riboswitch. *Nat Chem Biol, 13*(9), 968-974. doi:10.1038/nchembio.2427


A
EQ, 0mM ZMP
ZMP Responsive nts
No Change
Transcript Length
Nucleotide
ρ (Reactivity)
EQ, 1 mM ZMP
ZMP Binds
B
0 mM Intermediate Hairptin
IH1
C
0 mM, Pre-PK
P1
P2
D
0 mM 'Apo'
PK
P3
1 mM Intermediate Hairptin
1 mM, Pre-PK
1 mM, ZMP-Bound
E
PK Folding, EQ
A17
G25
A40
0 mM
1 mM
P1 Folds
PK Folds
ZMP Binds
PK Broken
ρ (Reactivity)
Transcript Length

<u>Figure 2 supplement 1.</u> SHAPE probing of equilibrated *Cbe pfl* ZTP riboswitch intermediates.

(A) Equilibrium refolded SHAPE-Seq reactivity matrix for the *Clostridium beijerinckii pfl* riboswitch with 0 mM and 1 mM ZMP. ZMP-responsive nts are highlighted in reactivity matrix cut-outs. The absence of data for transcripts 88 and 110 is due to ambiguous alignment of 3' ends.

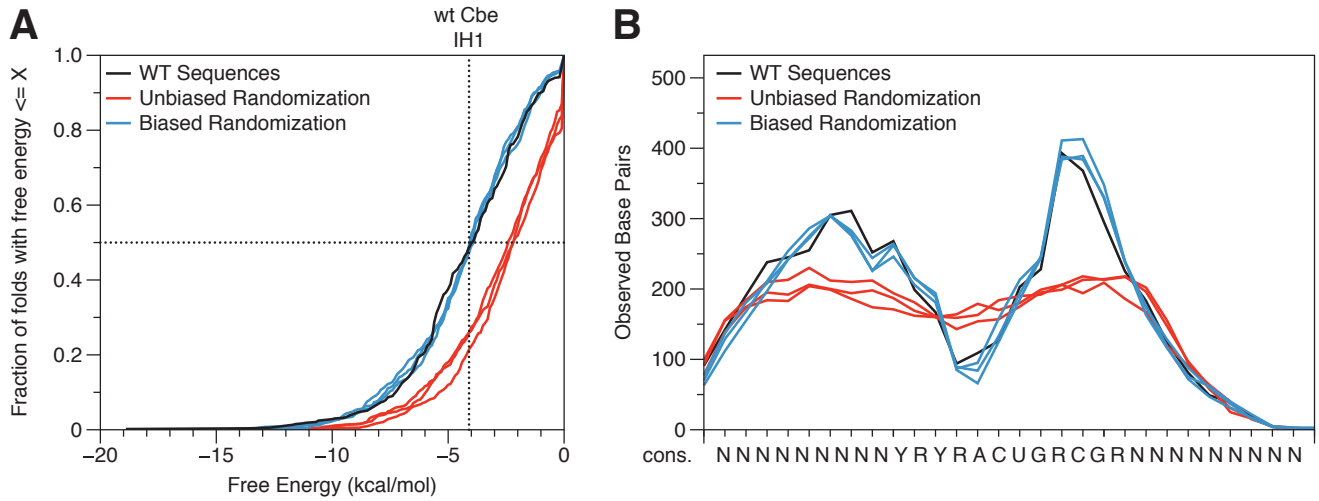
(B) Intermediate hairpin (IH1) secondary structure identified by manual analysis of reactivity values and refined with minimum free energy structure predictions of the leader region. Nucleotides are colored by the reactivity of transcript length 46 from (A). **(C)** Secondary structure of P1 and the linker region based, on the *pfl* RNA motif structure (Kim et al., 2015). Nucleotides are colored by the reactivity of transcript length 87 from (A).

(D) Apo and ZMP-bound secondary structure as determined by manual SHAPE-refinement of the consensus structure (Apo) (Kim et al., 2015) or by the ZTP aptamer crystal structures (Jones & Ferre-D'Amare, 2015; Ren et al., 2015; Trausch et al., 2015), respectively. Nucleotides are colored by the reactivity of transcript length 106 from (A).

(E) ZTP riboswitch folding as depicted by reactivity trajectories for nucleotides A17 (within IH1 loop or P1 stem, dotted lines), G25 (within pseudoknot, solid lines), and A40 (within P2 loop, dashed lines) from (A). Trajectories are shown for both the absence (purple) and presence (green) of 1mM ZMP. Notable structural changes are identified by vertical dashed lines and the position of critical intermediate structures in (B-D) are shown by vertical dotted lines. Shaded boxes indicate the regions of poor alignment described in (A). Note that the annotated folding events are shifted to transcript lengths ~14 nts earlier relative to cotranscriptionally folded and probed transcripts in Figure 2 due to the absence of RNA polymerase.

Results are for one experiment.

Figure 2 supplement 1-source data 1. Source data for Figure 2 supplement 1 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).



<u>Figure 2 supplement 2</u>. Analysis of ZTP riboswitch sequences for intermediate hairpin formation.

(A) Cumulative distribution plot of the minimum free energy (ΔG) predicted for ZTP riboswitch sequences that correspond to the *pfl* riboswitch intermediate hairpin structure. Sequences comprise the first nucleotide in the multiple sequence alignment (9-10 nts upstream of J1/2) through the last unstructured nt within J1/2. The analysis of several sequence pools is presented: (black) 532 ZTP riboswitches from fully sequenced bacterial genomes (Kim et al., 2015); (red) unbiased randomized sequences generated with an equal probability for observing each nucleotide at each position; (blue) sequences randomized by the natural nucleotide frequency at each position for the 532 sampled ZTP riboswitches (Kim et al., 2015). All randomized data sets match the length distribution observed in the wt ZTP riboswitch set.

(B) The number of base pairs predicted at each position as observed in the structures predicted in (A).

Figure 2 supplement 1-source data 2. Source data for Figure 2 supplement 2 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).

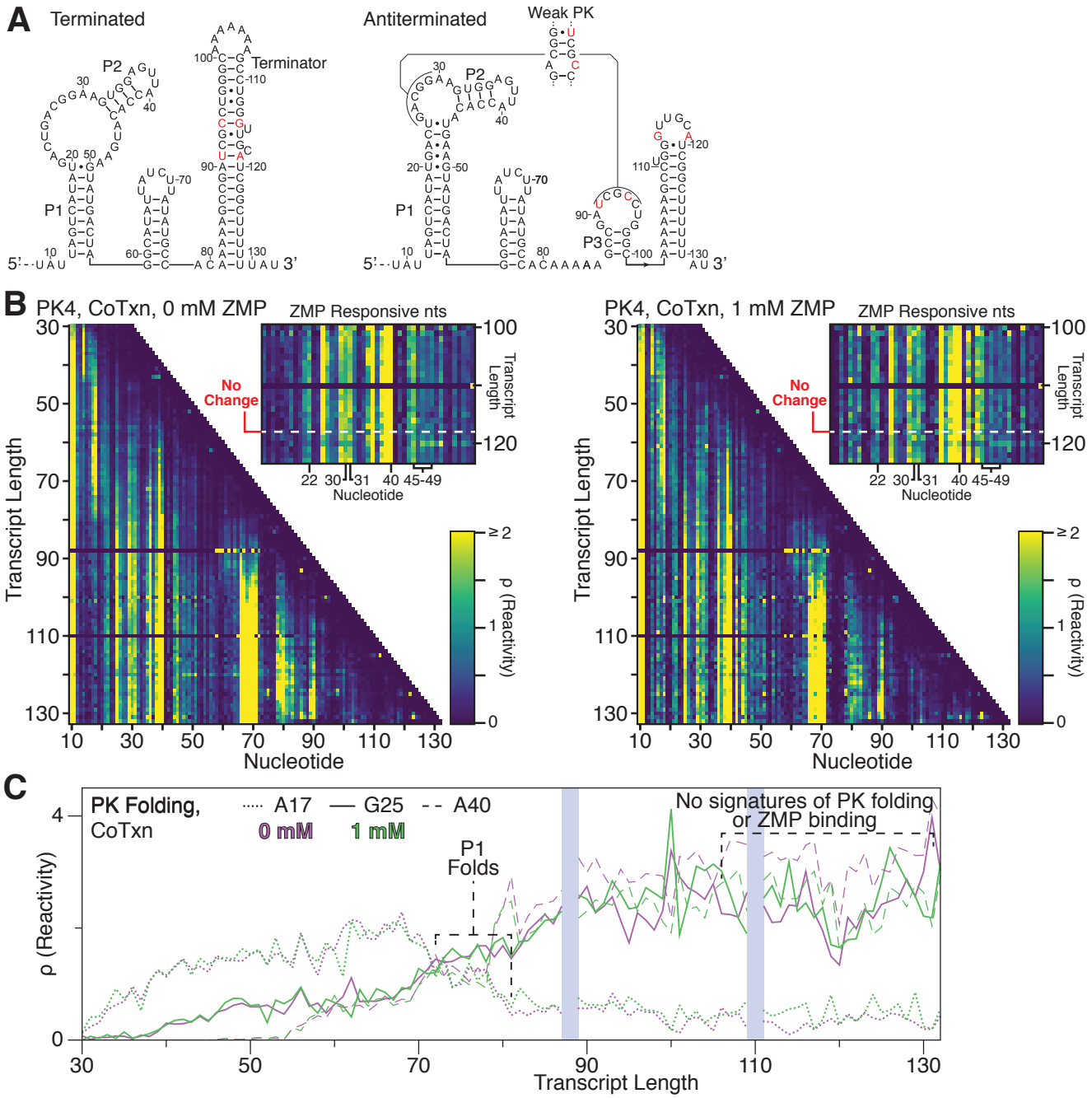


Figure 2 supplement 3. Cotranscriptional SHAPE probing of disrupted pseudoknot mutant.

(A) *Clostridium beijerinckii pfl* riboswitch terminated and antiterminated secondary structures shown with PK4 (C91U,U94C,A114G,G119A) mutations. The base pair configuration of the mutated pseudoknot is shown.

(B) Cotranscriptional SHAPE-Seq reactivity matrix for the *Clostridium beijerinckii* riboswitch PK4 mutant with 0 mM and 1 mM ZMP. ZMP-responsive nts are highlighted in reactivity matrix cut-outs. The absence of data for transcripts 88 and 110 is due to ambiguous alignment of 3' ends.

(C) ZTP riboswitch folding as depicted by reactivity trajectories for nucleotides A17 (within IH1 loop or P1 stem, dotted lines), G25 (within pseudoknot, solid lines), and A40 (within P2 loop, dashed lines) from (A). Trajectories

are shown for both the absence (purple) and presence (green) of 1mM ZMP. Notable structural changes are identified by vertical dashed lines. Shaded boxes indicate the regions of poor alignment described in (A). Results are for one experiment.

Figure 2 supplement 3-Source Data 1. Source data for Figure 2 supplement 3 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).

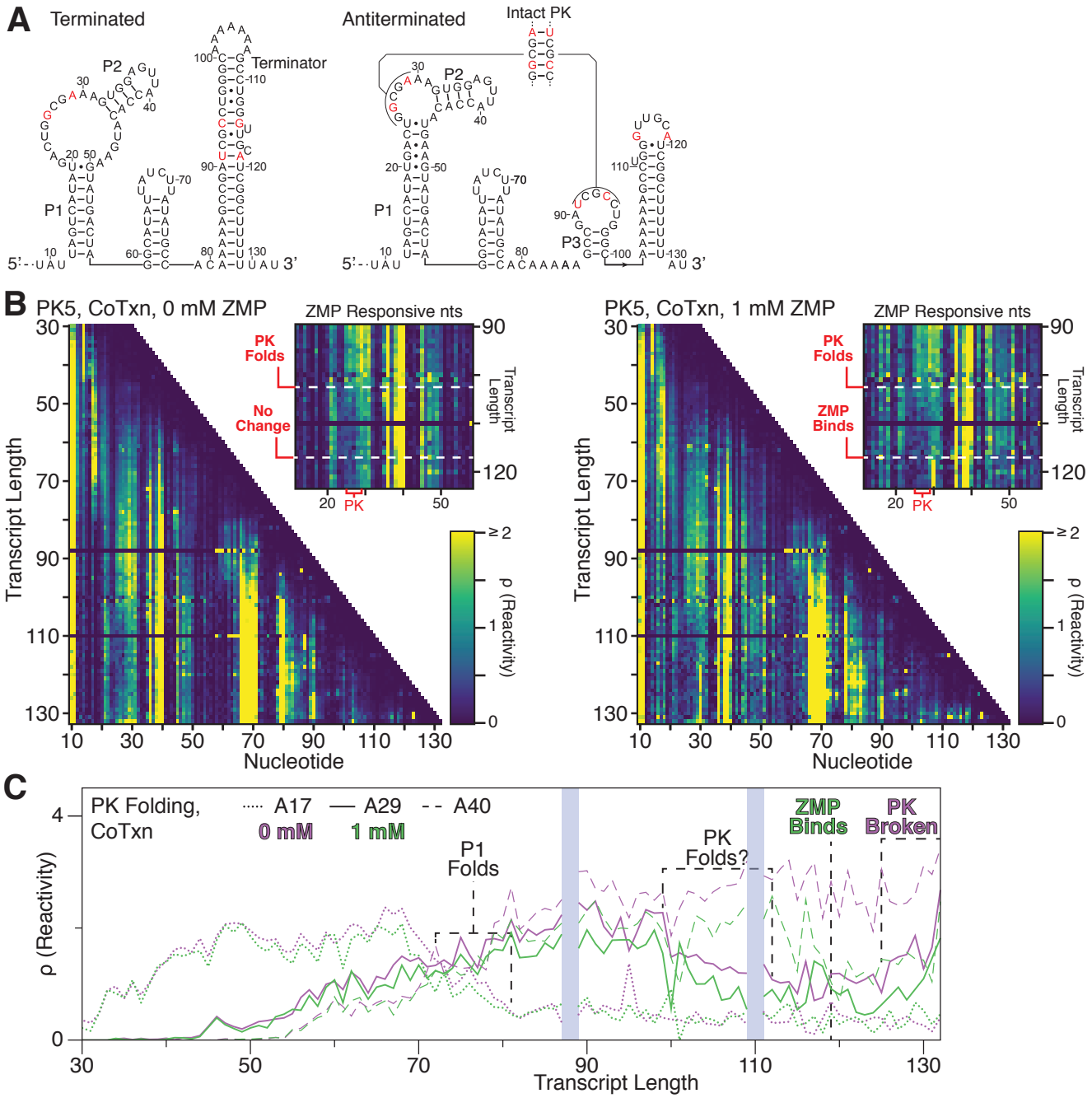


Figure 2 supplement 4. Cotranscriptional SHAPE probing of restored pseudoknot mutant.

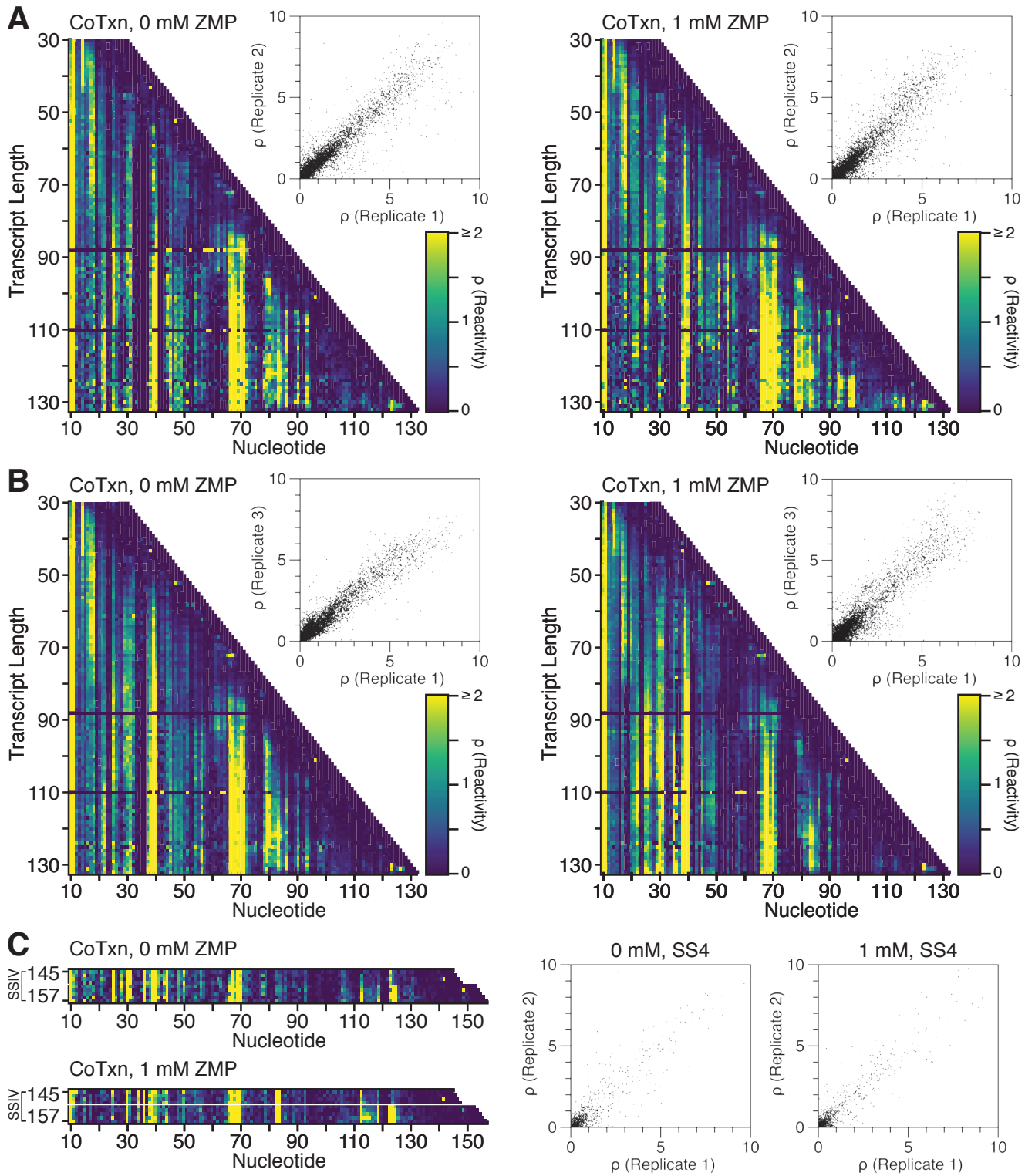
(A) *Clostridium beijerinckii* riboswitch terminated and antiterminated secondary structures shown with PK5 (A26G,G29A,C91U,U94C,A114G,G119A) mutations. The base pair configuration of the mutated pseudoknot is shown.

(B) Cotranscriptional SHAPE-Seq reactivity matrix for the *Clostridium beijerinckii* riboswitch PK5 mutant with 0 mM and 1 mM ZMP. ZMP-responsive nts are highlighted in reactivity matrix cut-outs. The absence of data for transcripts 88 and 110 is due to ambiguous alignment of 3' ends.

(C) ZTP riboswitch folding as depicted by reactivity trajectories for nucleotides A17 (within IH1 loop or P1 stem, dotted lines), A29 (within pseudoknot, solid lines), and A40 (within P2 loop, dashed lines) from (A). For PK5, pseudoknot nucleotides from 25-29 were more evenly reactive than the wt sequence and A29 is shown as a representative for clarity. Trajectories are shown for both the absence (purple) and presence (green) of 1mM ZMP. Notable structural changes are identified by vertical dashed lines. Shaded boxes indicate the regions of poor alignment described in (A).

Results are for one experiment.

Figure 2 supplement 4-Source Data 1. Source data for Figure 2 supplement 4 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).



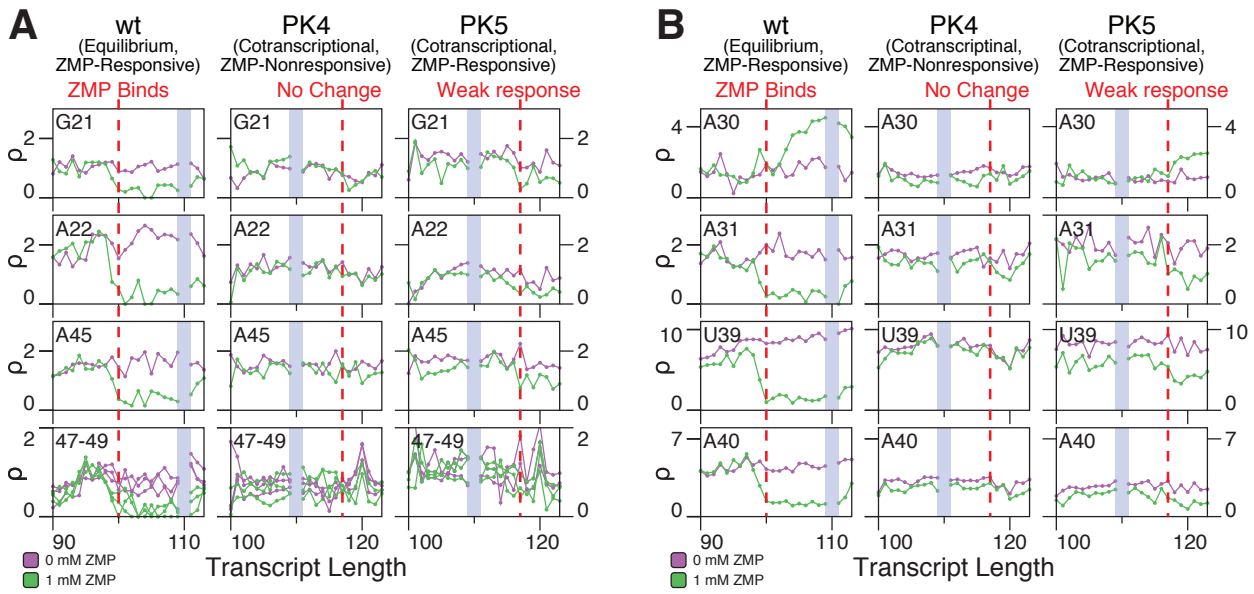
<u>Figure 2 supplement 5.</u> Cotranscriptional SHAPE probing of the *Cbe pfl* ZTP riboswitch Replicate Data.

(A) Cotranscriptional SHAPE-Seq reactivity matrix for the *Clostridium beijerinckii pfl* ZTP riboswitch with 0 mM and 1 mM ZMP; Replicate 2. The absence of data for transcripts 88 and 110 is due to ambiguous alignment of 3' ends. Reactivity values from replicate 2 transcript lengths 30 to 132 are plotted against the corresponding values from replicate 1 for each condition.

(B) As in (A), but for replicate 3.

(C) Cotranscriptional SHAPE-Seq reactivity matrix for the *Clostridium beijerinckii pfl* ZTP riboswitch with 0 mM and 1 mM ZMP targeting transcripts 145-148 and 153-157 using terminal biotin roadblocks and Superscript IV reverse transcriptase (RT); Replicate 2. Reactivity values from replicate 2 transcript lengths 30 to 132 are plotted against the corresponding values from replicate 1 for each condition.

Figure 2 supplement 5-Source Data 1. Source data for Figure 2 supplement 5 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).

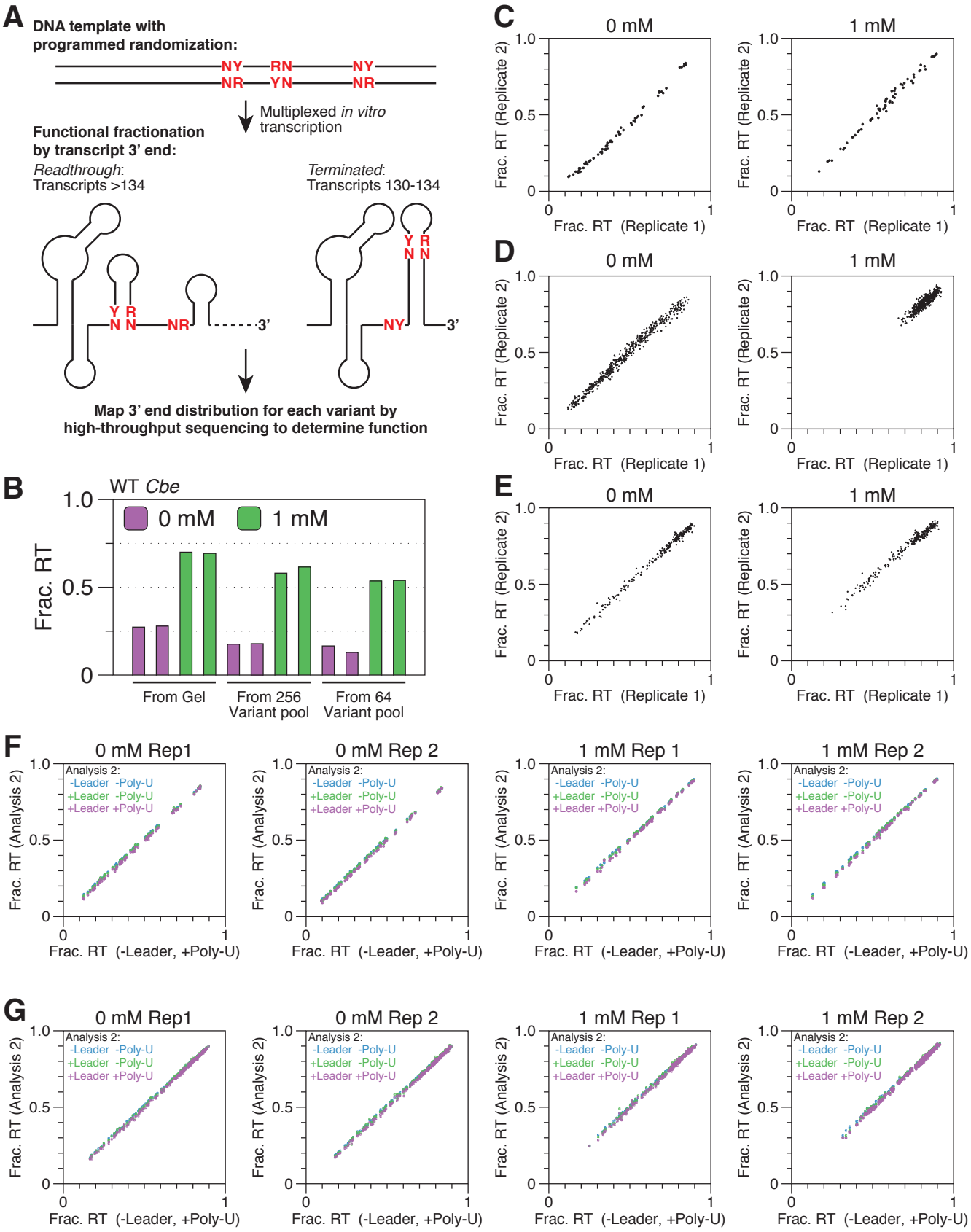


<u>Figure 3 supplement 1</u>. SHAPE reactivity of ZMP-responsive nucleotides in equilibrated wt and cotranscriptionally folded pseudoknot mutants.

ZMP-responsive nucleotides within the *Cbe pfl* ZTP riboswitch, including **(A)** non-canonical P1 base pairs and **(B)** the nucleotides that interact with the pseudoknot as defined in Figure 3. SHAPE-Seq reactivity traces across transcript lengths are shown in the absence (purple) and presence (green) of 1mM ZMP. Wt equilibrium refolded data are from Figure 2 supplement 1 and PK4/PK5 cotranscriptionally folded data are from Figure 2 supplements 3 and 4. The vertical red dashed lines indicate the transcript length at which ZMP binding is inferred to occur based on bifurcation of the WT trajectories.

Results are for one experiment.

Figure 3 supplement 1-source data 1. Source data for Figure 3 supplement 1 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).


A
DNA template with programmed randomization:
NY RN NY
NR YN NR
Multiplexed in vitro transcription
Functional fractionation by transcript 3' end:
Readthrough: Transcripts >134
Terminated: Transcripts 130-134
Y R N N NR 3'
Y R N N NY 3'
Map 3' end distribution for each variant by high-throughput sequencing to determine function
B
WT Cbe
0 mM
1 mM
Frac. RT
From Gel
From 256 Variant pool
From 64 Variant pool
C
D
E
0 mM
1 mM
Frac. RT (Replicate 2)
Frac. RT (Replicate 1)
F
G
0 mM Rep1
0 mM Rep 2
1 mM Rep 1
1 mM Rep 2
Analysis 2:
-Leader -Poly-U
+Leader -Poly-U
+Leader +Poly-U
Frac. RT (Analysis 2)
Frac. RT (-Leader, +Poly-U)

<u>Figure 4 supplement 1.</u> Targeted *in vitro* transcription and data analysis controls.

(A) Overview of combinatorial mutagenesis experiment. A linear DNA template library is constructed using a long synthetic oligonucleotide with randomization at specified positions. Multiplexed *in vitro* transcription is performed using this DNA template library to produce a distribution of 3' ends for each variant within the library. High-throughput sequencing is then used to measure this distribution for every variant to determine its function.

(B) Fraction readthrough for the wt *Cbe* ZTP riboswitch with 0 mM and 1 mM as measured by targeted *in vitro* transcription and gel electrophoresis (data from Fig. 1C) or by high-throughput sequencing as part of a combinatorial mutagenesis library with 256 (data from experiment in Figure 6) or 64 (data from experiment in Figure 4) variants.

(C) Plot of fraction readthrough values for all aptamer/terminator overlap combinatorial mutagenesis variants from replicate 1 against replicate 2 for 0 mM and 1 mM ZMP conditions (data from Figure 4).

(D) Plot of fraction readthrough values for all P3 stem combinatorial mutagenesis variants from replicate 1 against replicate 2 for 0 mM and 1 mM ZMP conditions (data from Figure 5).

(D) Plot of fraction readthrough values for all pseudoknot contact combinatorial mutagenesis variants from replicate 1 against replicate 2 for 0 mM and 1 mM ZMP conditions (data from Figure 6).

(F) Analysis controls in which sequencing alignment omitted the ZTP riboswitch 5' leader, poly-U tract, or both. Each analysis control plotted against alignment without the 5' leader and with the poly-U tract. Replicate 1 data are from Figure 4 and Replicate 2 data are from Figure 4 Supplement 2.

(G) As in (F) but Replicate 1 Data are from Figure 6 and Replicate 2 data are from Figure 6 Supplement 4.

Figure 4 Supplement 1-source data 1. Source data for Figure 4 Supplement 1 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).

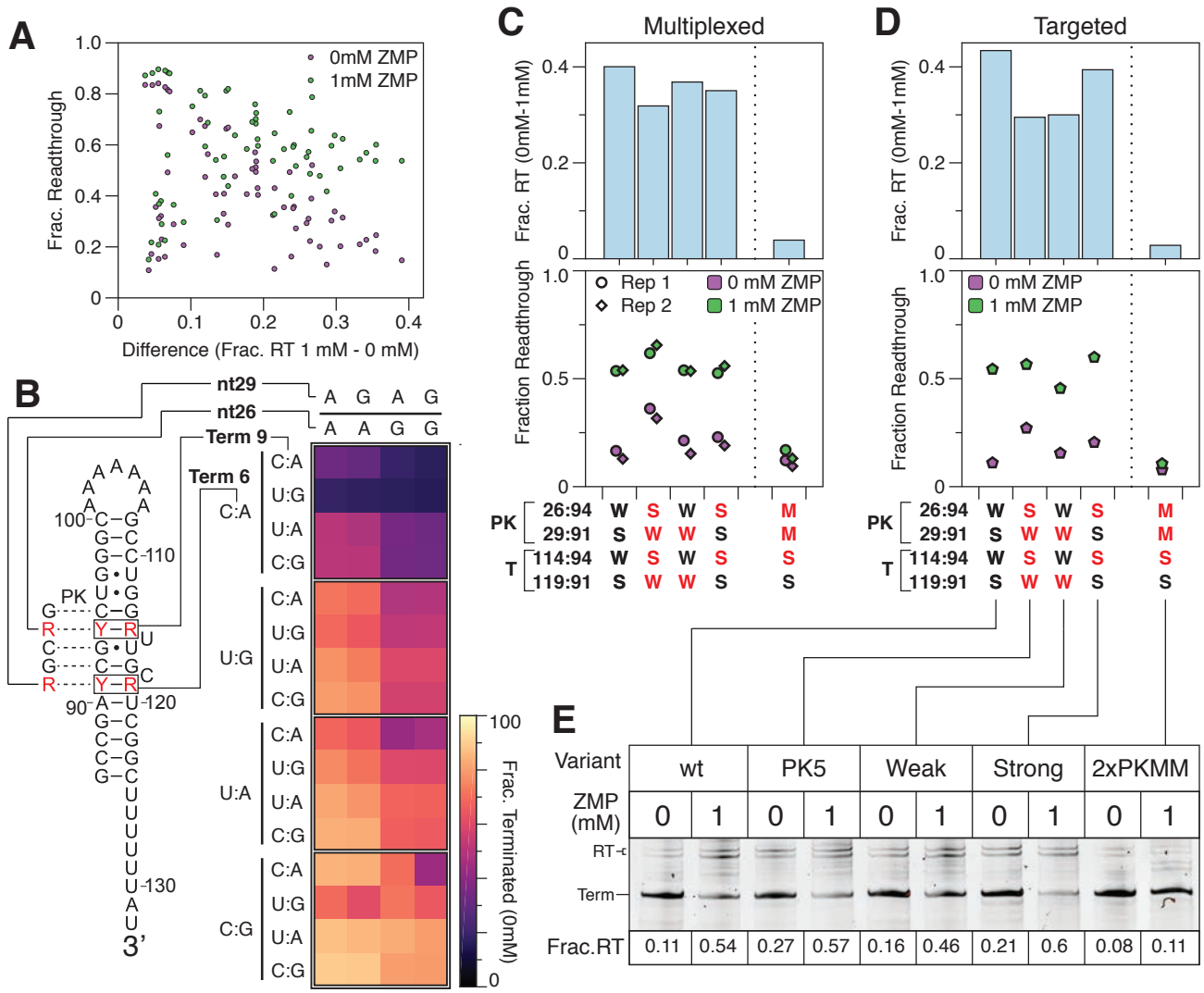


Figure 4 supplement 2. Complete aptamer/terminator overlap combinatorial mutagenesis data.

(A) Plot of fraction readthrough of aptamer/terminator overlap combinatorial mutagenesis mutants from Figure 4A as measured in the absence (purple) and presence (green) of 1 mM ZMP ordered by the difference in fraction readthrough observed in 1 mM and 0 mM ZMP conditions.

(B) Terminator efficiency (1 – fraction readthrough) of aptamer/terminator overlap combinatorial mutagenesis mutants from Figure 4A in the absence of ZMP. Mutants are grouped vertically by the inferred stability of terminator base pairs 91:119 (Term 6) and 94:114 (Term 9) and horizontally by the inferred stability of pseudoknot base pairs 26:94 and 29:91.

(C) Fraction readthrough for mutants with complete pseudoknot and terminator base pair matches and for a mutant with a complete terminator match and pseudoknot mismatch. The average difference in fraction readthrough (1 mM – 0 mM) is shown above each variant. Variant pairing patterns in the pseudoknot (PK) and

terminator (T) are annotated as weak (W, A-U), strong (S, G-C), and mismatch (M, A-C). Red annotations indicate deviations from the wild-type pairing pattern.

(D) Fraction readthrough for mutants from (C) as measured by targeted *in vitro* transcription and gel electorophresis.

(E) *in vitro* transcription gel for measurements in (D).

n=2 independent biological replicates are annotated as 'Rep 1' and 'Rep 2' in panel C. Panels A and B are the average of n=2 independent biological replicates. Individual replicate values are compared in Figure 4-figure supplement 1C. Panels D and E are from one experiment.

Figure 4 supplement 2-source data 1. Source data for Figure 4 supplement 2 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).

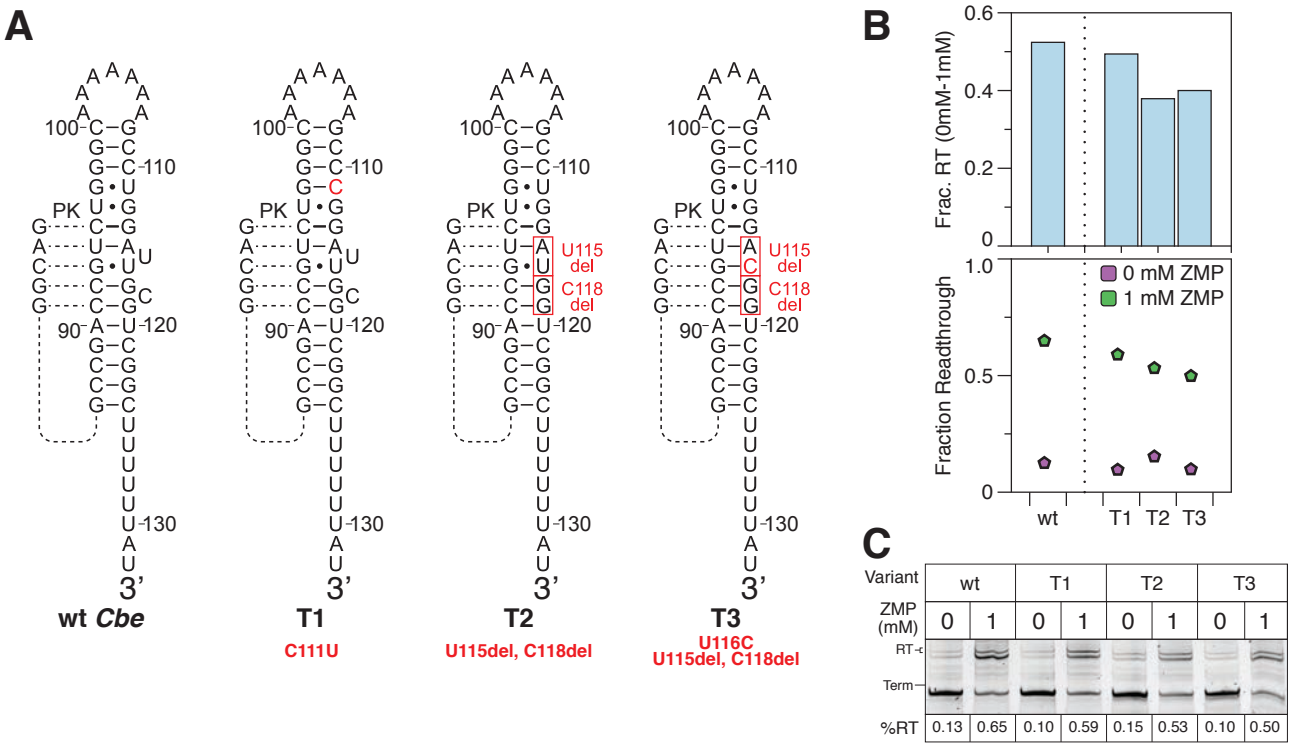


Figure 4 supplement 3. ***in vitro*** **transcription of** ***Cbe*** **ZTP riboswitch terminator variants**

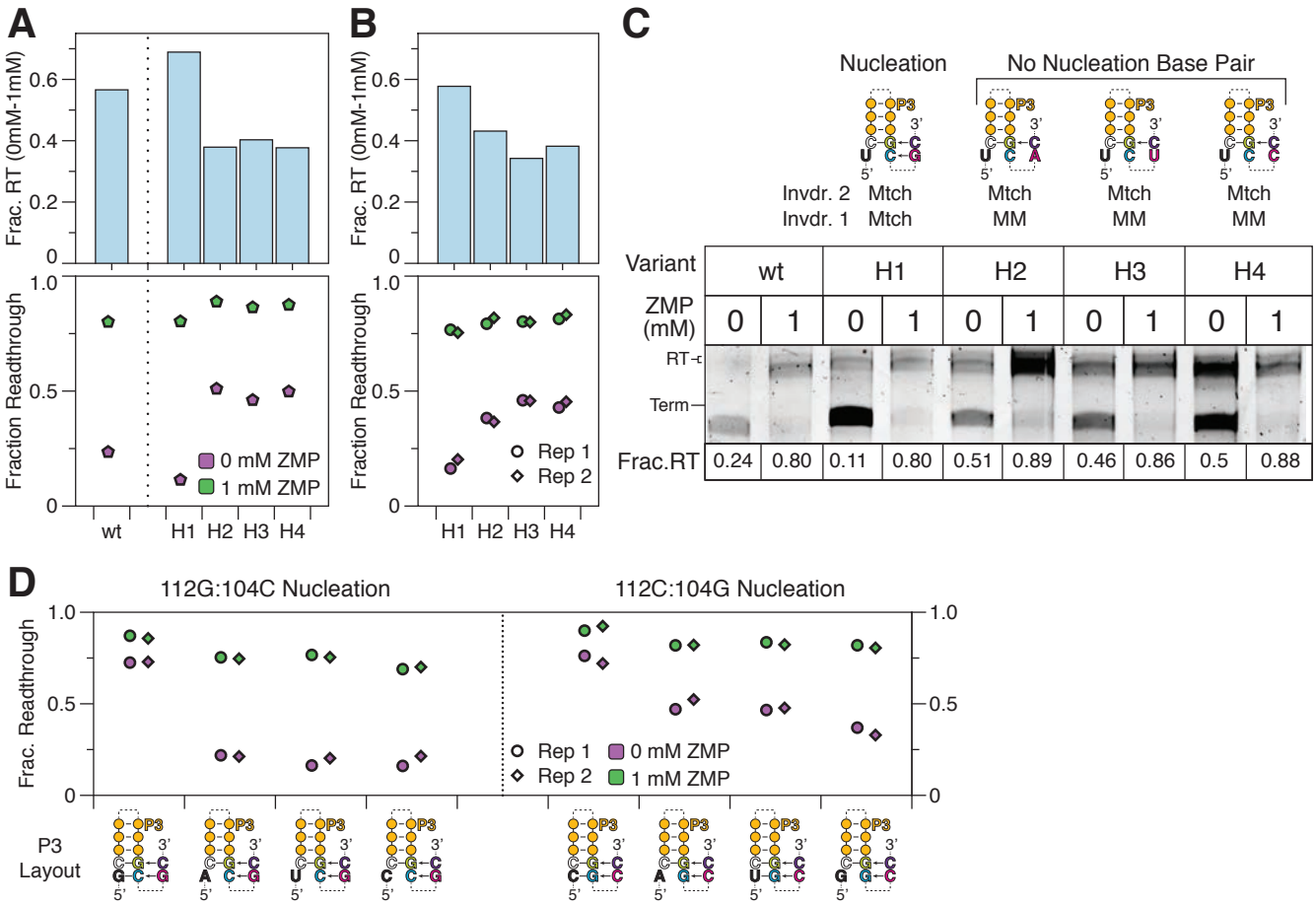
(A) Expected secondary structure of intrinsic terminator hairpin for mutants analyzed in (B). The position of pseudoknot base pairs that compete with the formation of the 3' terminator stem is shown.

(B) Fraction readthrough for terminator mutants shown in (A) as measured in the absence (purple) and presence (green) of 1 mM ZMP. The difference in fraction readthrough (1 mM – 0 mM) is shown above each variant.

(C) *in vitro* transcription gel for measurements in (B).

Results are from one experiment.

Figure 4 Supplement 3-source data 1. Source data for Figure 4 supplement 3 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).



<u>Figure 5 Supplement 1.</u> *in vitro* transcription of *Cbe pfl* ZTP riboswitch terminator nucleation variants

(A) Targeted *in vitro* transcription of wt *Cbe pfl* riboswitch and terminator nucleation variants. Fraction readthrough for variants as measured in the absence (purple) and presence (green) of 1 mM ZMP. The difference in fraction readthrough (1 mM – 0 mM) is shown above each variant. The configuration for terminator strand displacement of P3 for each variant is shown in (C).

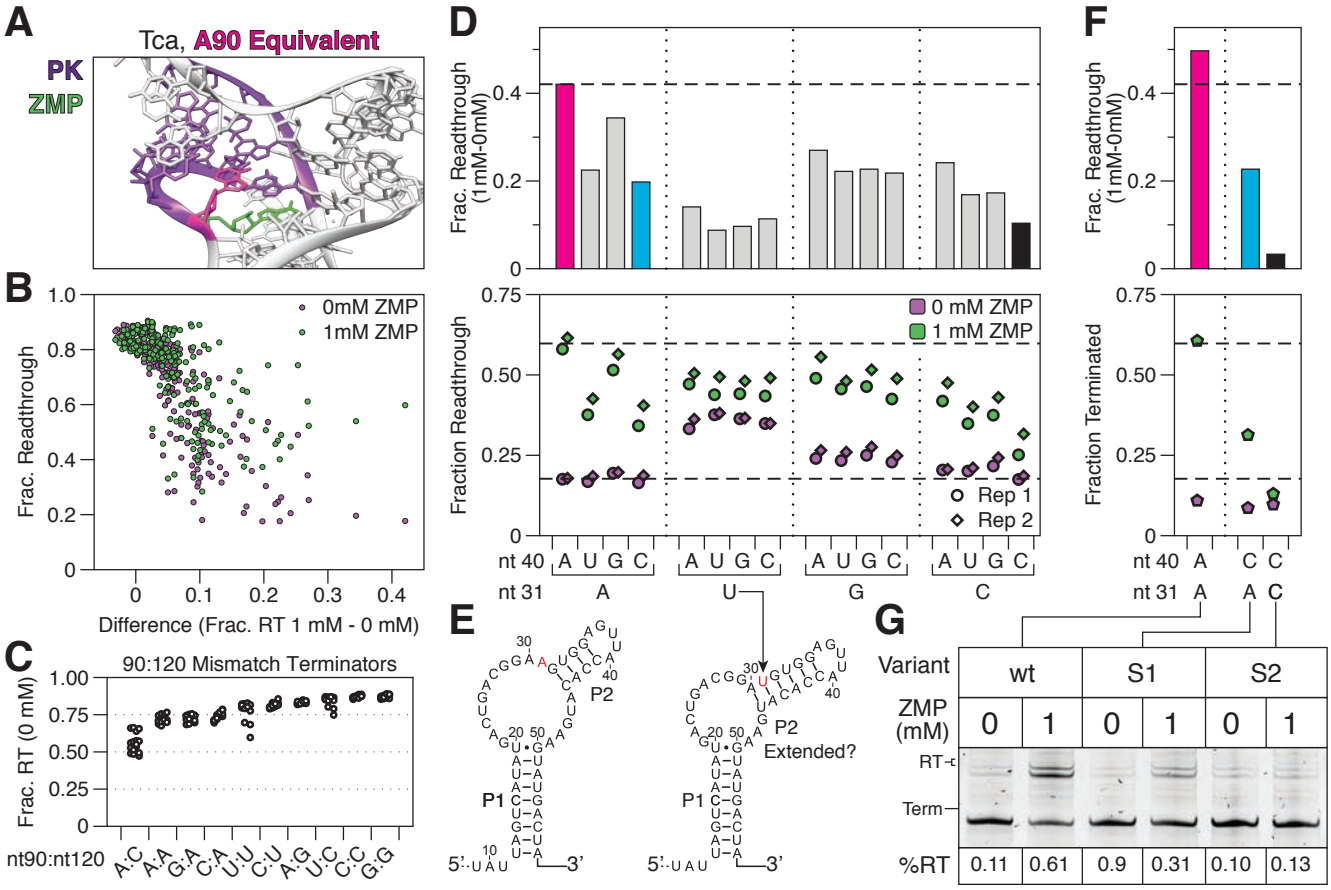
(B) Fraction readthrough for variants in (A) as measured by combinatorial mutagenesis.

(C) *In* vitro transcription gel for the measurements in (A). The configuration for terminator strand displacement of P3 for each variant is shown as in Figure 5.

Results in (A) are from one experiment. Results in (B) are from n=2 independent biological replicates; the difference in fraction readthrough (0mM-1mM) is the average of these replicates.

(D) Fraction readthrough for select variants with terminator hairpin nucleation by 112G:104C or 112:C104G base pairs.

Figure 5 source data 1. Source data for Figure 5 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).



<u>Figure 6 supplement 1.</u> Complete combinatorial mutagenesis data and targeted *in vitro* transcription analysis of pseudoknot contact variants.

(A) The crystal structure of the *Thermosinus carboxydivorans* ZMP riboswitch (PDB: 4ZNP) (Ren et al., 2015) is highlighted to show the interaction between the nucleotide equivalent to *Cbe pfl* A90 (magenta) interacting with the Hoogsteen edge of a pseudoknot nucleotide (purple).

(B) Plot of fraction readthrough of combinatorial mutagenesis mutants from Figure 6A as measured in the absence (purple) and presence (green) of 1 mM ZMP ordered by the difference in fraction readthrough observed in 1 mM and 0 mM ZMP conditions.

(C) Fraction readthrough in the absence of ZMP for all variants from Figure 6A with a mismatch between nucleotides 90 and 120.

(D) Fraction readthrough for nucleotide 31 and 40 mutants with the wt terminator (90A, 120U). Dashed lines indicate fraction readthrough observed for the wt *Cbe* ZTP riboswitch in each condition. The average difference in fraction readthrough (1 mM – 0 mM is shown as a bar plot). Bar colors indicate mutants that were tested in targeted *in vitro* transcription expeirments in (F).

(E) Secondary structure of the wt *Cbe* ZTP aptamer and an extended P2 variation that could occur for A31U mutants.

(F) Targeted *in vitro* transcription analysis of select mutants from (D).

(G) *In vitro* transcription gel for measurements in (F).

n=2 independent biological replicates are annotated as 'Rep 1' and 'Rep 2' in panel D. Panels B and C are the average of n=2 independent biological replicates. Individual replicate values are compared in Figure 4-figure supplement 1. Panels F and G are from one experiment.

Figure 6 supplement 1-source data 1. Source data for Figure 6 supplement 1 are available in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N20F4H).

Supplementary Materials for

A mechanism for ligand gated strand displacement in ZTP riboswitch transcription regulation

Eric J. Strobel[1]*, Luyi Cheng[2,4], Katherine E. Berman[2,4], Paul D. Carlson[3,4], Julius B. Lucks[1,4]*

Affiliations:
[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA
[2]Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL 60208, USA
[3]Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, 14853, USA
[4]Center for Synthetic Biology, Northwestern University, Evanston, IL, 20208

*Correspondence to: Email: eric.strobel@northwestern.edu, jblucks@northwestern.edu

This pdf file includes:

Supplementary Tables S1 to S6

Supplementary Table 1. Sequencing Read Archive (SRA) deposition table.

All primary sequencing data generated in this work are freely available from the Sequencing Read Archive (http://www.ncbi.nlm.nih.gov/sra), accessible via the BioProject accession number PRJNA510362 or using the individual accession numbers below.

SRA Accession	RNA	Experiment	Figure(s)
SAMN10607306	ZTP riboswitch, wt (rep.1)	0 mM ZMP, cotranscriptional	Figs. 2, 3
SAMN10607307	ZTP riboswitch, wt (rep. 2)	0 mM ZMP, cotranscriptional	Fig. 2 Supp. 5
SAMN10607308	ZTP riboswitch, wt (rep. 3)	0 mM ZMP, cotranscriptional	Fig. 2 Supp. 5
SAMN10607309	ZTP riboswitch, wt (rep. 1)	1 mM ZMP, cotranscriptional	Figs. 2, 3
SAMN10607310	ZTP riboswitch, wt (rep. 2)	1 mM ZMP, cotranscriptional	Fig. 2 Supp. 5
SAMN10607311	ZTP riboswitch, wt (rep. 3)	1 mM ZMP, cotranscriptional	Fig. 2 Supp. 5
SAMN10607312	ZTP riboswitch, wt (SSIV rep. 1)	0 mM ZMP, cotrans, SSIV RT	Fig 2. Supp. 5
SAMN10607313	ZTP riboswitch, wt (SSIV rep. 2)	0 mM ZMP, cotrans, SSIV RT	Fig. 2
SAMN10607314	ZTP riboswitch, wt (SSIV rep. 1)	1 mM ZMP, cotrans, SSIV RT	Fig 2. Supp. 5
SAMN10607315	ZTP riboswitch, wt (SSIV rep. 2)	1 mM ZMP, cotrans, SSIV RT	Fig. 2
SAMN10607316	ZTP riboswitch, wt	0 mM ZMP, equilibrium	Fig. 2 Supp. 1. Fig. 3 Supp. 1.
SAMN10607317	ZTP riboswitch, wt	1 mM ZMP, equilibrium	Fig. 2 Supp. 1. Fig. 3 Supp. 1.
SAMN10607318	ZTP riboswitch, PK4 Mutant	0 mM ZMP, cotranscriptional	Fig. 2 Supp. 3
SAMN10607319	ZTP riboswitch, PK4 Mutant	1 mM ZMP, cotranscriptional	Fig. 2 Supp. 3
SAMN10607320	ZTP riboswitch, PK5 Mutant	0 mM ZMP, cotranscriptional	Fig. 2 Supp. 4
SAMN10607321	ZTP riboswitch, PK5 Mutant	1 mM ZMP, cotranscriptional	Fig. 2 Supp. 4
SAMN10607322	ZTP RndPK (rep. 1)	0 mM ZMP	Fig. 4 Fig. 4 Supp. 1 Fig. 4 Supp. 2
SAMN10607323	ZTP RndPK (rep. 2)	0 mM ZMP	Fig. 4 Fig. 4 Supp. 1 Fig. 4 Supp. 2
SAMN10607324	ZTP RndPK (rep. 1)	1 mM ZMP	Fig. 4 Fig. 4 Supp. 1 Fig. 4 Supp. 2
SAMN10607325	ZTP RndPK (rep. 2)	1 mM ZMP	Fig. 4 Fig. 4 Supp. 1 Fig. 4 Supp. 2
SAMN10607326	ZTP RndP3v2 (rep. 1)	0 mM ZMP	Fig. 5 Fig. 5 Supp. 1 Fig. 4 Supp. 1
SAMN10607327	ZTP RndP3v2 (rep. 2)	0 mM ZMP	Fig. 5 Fig. 5 Supp. 1 Fig. 4 Supp. 1
SAMN10607328	ZTP RndP3v2 (rep. 1)	1 mM ZMP	Fig. 5 Fig. 5 Supp. 1 Fig. 4 Supp. 1
SAMN10607329	ZTP RndP3v2 (rep. 2)	1 mM ZMP	Fig. 5 Fig. 5 Supp. 1
SAMN10607330	ZTP RndPKSTB (rep. 1)	0 mM ZMP	Fig. 6 Fig. 6 Supp. 1 Fig. 4 Supp. 1
SAMN10607331	ZTP RndPKSTB (rep. 2)	0 mM ZMP	Fig. 6 Fig. 6 Supp. 1 Fig. 4 Supp. 1
SAMN10607332	ZTP RndPKSTB (rep. 1)	1 mM ZMP	Fig. 6 Fig. 6 Supp. 1 Fig. 4 Supp. 1
SAMN10607333	ZTP RndPKSTB (rep. 2)	1 mM ZMP	Fig. 6 Fig. 6 Supp. 1 Fig. 4 Supp.1

Supplementary Table 2. RMDB data deposition table.

SHAPE-Seq reactivity spectra generated in this work are freely available from the RNA Mapping Database (RMDB) (http://rmdb.stanford.edu/repository/) (Cordero et al., 2012), accessible using the RMDB ID numbers indicated in the table below.

RMDB Accession	RNA	Experiment	Figure(s)
ZTPRSW_BZCN_0001	ZTP riboswitch, wt (rep.1)	0 mM ZMP, cotranscriptional	Figs. 2, 3
ZTPRSW_BZCN_0002	ZTP riboswitch, wt (rep. 2)	0 mM ZMP, cotranscriptional	Fig. 2 Supp. 5
ZTPRSW_BZCN_0003	ZTP riboswitch, wt (rep. 3)	0 mM ZMP, cotranscriptional	Fig. 2 Supp. 5
ZTPRSW_BZCN_0004	ZTP riboswitch, wt (rep. 1)	1 mM ZMP, cotranscriptional	Figs. 2, 3
ZTPRSW_BZCN_0005	ZTP riboswitch, wt (rep. 2)	1 mM ZMP, cotranscriptional	Fig. 2 Supp. 5
ZTPRSW_BZCN_0006	ZTP riboswitch, wt (rep. 3)	1 mM ZMP, cotranscriptional	Fig. 2 Supp. 5
ZTPRSW_BZCN_0011	ZTP riboswitch, wt	0 mM ZMP, equilibrium	Fig 2. Supp. 5
ZTPRSW_BZCN_0012	ZTP riboswitch, wt	1 mM ZMP, equilibrium	Fig. 2
ZTPRSW_BZCN_0007	ZTP riboswitch, wt (SSIV rep. 1)	0 mM ZMP, cotranscriptional, SSIV RT	Fig 2. Supp. 5
ZTPRSW_BZCN_0008	ZTP riboswitch, wt (SSIV rep. 2)	0 mM ZMP, cotranscriptional, SSIV RT	Fig. 2
ZTPRSW_BZCN_0009	ZTP riboswitch, wt (SSIV rep. 1)	1 mM ZMP, cotranscriptional, SSIV RT	Fig. 2 Supp. 1. Fig. 3 Supp. 1.
ZTPRSW_BZCN_0010	ZTP riboswitch, wt (SSIV rep. 2)	1 mM ZMP, cotranscriptional, SSIV RT	Fig. 2 Supp. 1. Fig. 3 Supp. 1.
ZTPRSW_BZCN_0013	ZTP riboswitch, PK4 Mutant	0 mM ZMP, cotranscriptional	Fig. 2 Supp. 3
ZTPRSW_BZCN_0014	ZTP riboswitch, PK4 Mutant	1 mM ZMP, cotranscriptional	Fig. 2 Supp. 3
ZTPRSW_BZCN_0015	ZTP riboswitch, PK5 Mutant	0 mM ZMP, cotranscriptional	Fig. 2 Supp. 4
ZTPRSW_BZCN_0016	ZTP riboswitch, PK5 Mutant	1 mM ZMP, cotranscriptional	Fig. 2 Supp. 4

Supplementary Table 3. Sequences used for in vitro transcription templates.

Summary of the sequences used for generating *in vitro* transcription templates. The ZTP riboswitch contains a two nucleotide insertion at the 5' end of the template relative to the wild type *Clostridium beijerinckii* sequence. Templates were extended with a ribosome binding site (RBS) and a portion of the superfolder GFP (SFGFP) coding sequence, listed as 'trailing sequence' below. Promoter sequences are blue. Mutation positions are highlighted with red. Deletion positions are shown as "-".

Description	Sequence
Promoter	GCTTGATTCTAAAGATCTTTGACAGCTAGCTCAGTCCTAGGTATAATACTAGT
C. beijericnckii ZTP *pfl* riboswitch	ATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGACCGTCTGGGCAAAAAAAGCCTGGATTGCGTCGGCTTTTTTAT
Cbe ZTP riboswitch PK4	ATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGATCGCCTGGGCAAAAAAAGCCTGGGTTGCATCGGCTTTTTTAT
Cbe ZTP riboswitch PK5	ATATTAGATATTAGTCATATGACTGGCGAAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGATCGCCTGGGCAAAAAAAGCCTGGGTTGCATCGGCTTTTTTAT
Cbe ZTP riboswitch PKAT	ATATTAGATATTAGTCATATGACTGACGAAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGATCGTCTGGGCAAAAAAAGCCTGGATTGCATCGGCTTTTTTAT
Cbe ZTP riboswitch PKGC	ATATTAGATATTAGTCATATGACTGGCGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGACCGCCTGGGCAAAAAAAGCCTGGGTTGCGTCGGCTTTTTTAT
Cbe ZTP riboswitch PKMM	ATATTAGATATTAGTCATATGACTGACGAAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGACCGCCTGGGCAAAAAAAGCCTGGGTTGCGTCGGCTTTTTTAT
Cbe ZTP riboswitch T1	ATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGACCGTCTGGGCAAAAAAAGCCCGGATTGCGTCGGCTTTTTTAT
Cbe ZTP riboswitch T2	ATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGACCGTCTGGGCAAAAAAAGCCTGGA-TG-GTCGGCTTTTTTAT
Cbe ZTP riboswitch T3	ATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGACCGTCTGGGCAAAAAAAGCCTGGA-CG-GTCGGCTTTTTTAT
Cbe ZTP riboswitch H1	ATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAATCGCCGACCGTCTGGGCGCAAAAAAAGCGCCTGGATTGCGTCGGCTTTTTTAT
Cbe ZTP riboswitch H2	ATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAATCGCCGACCGTCTGGGCGCAAAAAAAACGCCTGGATTGCGTCGGCTTTTTTAT
Cbe ZTP riboswitch H3	ATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAATCGCCGACCGTCTGGGCGCAAAAAAATCGCCTGGATTGCGTCGGCTTTTTTAT
Cbe ZTP riboswitch H4	ATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAATCGCCGACCGTCTGGGCGCAAAAAAACCGCCTGGATTGCGTCGGCTTTTTTAT
Cbe ZTP riboswitch S1	ATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTCCCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGACCGTCTGGGCAAAAAAAGCCTGGATTGCGTCGGCTTTTTTAT
Cbe ZTP riboswitch S2	ATATTAGATATTAGTCATATGACTGACGGACGTGGAGTTCCCACATGAAGTATGACTAGGCATATTATCTTATATG CCACAAAAAGCCGACCGTCTGGGCAAAAAAAGCCTGGATTGCGTCGGCTTTTTTAT
Trailing Sequence	atggaaaaggaggaaggatctatgagcaaaggagaagaacttttcactggagttgtcccaattcttgttg

Supplementary Table S4. Oligonucleotides used for DNA template amplification

Below is a table of oligonucleotides used for the preparation of *in vitro* transcription DNA templates. Abbreviations within primer sequences are as follows: '/5Biosg/' is a 5' biotin moiety. These abbreviations were used for compatibility with the Integrated DNA Technologies ordering notation.

Description	Sequence	Purification	ID
Fwd. primer for DNA templates	GCTTCCGGCTGAATTCTAAAGATCT	None	A
Rev. primer DNA templates	/5Biosg/CAACAAGAATTGGGACAACTCCAGTG	None	B
Fwd. primer for randomly biotinylated DNA templates	CTAAAGATCTTTGACAGCTAGCTCAGTCCTAGGTAT AATACTAGT	None	C
Rev. primer for 155 nt ZTP terminal roadblock	/5BiosG/ATAGATCCTTCCTCCTTTTCCATATA	None	D
Rev. primer for 165 nt ZTP terminal roadblock	/5BiosG/TCCTTTGCTCATAGATCCTTCCTCC	None	E
Fwd. primer for initial Ultramer PCR	CTAAAGATCTTTGACAGCTAGCTCAGTCCTAGGTATAAT	HPLC	F
Fwd. primer for preparatory mutagenesis library PCR	CTAAAGATCTTTGACAGCTAGCTCAGTC	None	G

Supplementary Table S5. Ultramer Mutagenesis Oligonucleotides

Below is a table of Ultramer oligonucleotides used for the preparation of *in vitro* transcription DNA templates for combinatorial functional mutagenesis experiments. Uppercase sequence indicates the riboswitch. Lowercase sequence indicates flanking regions. IUPAC nucleotide notation is used. Insertion mutations are designated in the description by 'i' followed by the nucleotide number that that the insertion precedes; multiple insertions at a given position are designated by apostrophes. For example, NY insertions at nucleotide 86 are designated i86'N for the first insertion and i86''Y for the second insertion.

Description	Sequence	Purification	ID
Randomized Pseudoknot: 26R, 29R, 91Y, 94Y, 114R, 119R	ttgacagctagctcagtcctaggtataatactagtATATTAGATATTAGTCATATGACTGRCGRAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATGCCACAAAAAGCCGAYCGYCTGGGCAAAAAAAGCCTGGRTTGCRTCGGCTTTTTTATatggaaaaggaggaaggatctatgagcaaagga	PAGE Ultramer	H
Randomized P3 Insertions: i86'N, i86''Y i101'R, i101''N i107'N, i107''Y	acagctagctcagtcctaggtataatactagtATATTAGATATTAGTCATATGACTGACGGAAGTGGAGTTACCACATGAAGTATGACTAGGCATATTATCTTATATGCCACAAAAANYGCCGACCGTCTGGGCRNAAAAAAANYGCCTGGATTGCGTCGGCTTTTTTATatggaaaaggaggaaggatctatgagcaaa	PAGE Ultramer	I
Randomized Pseudoknot Contacting Nucleotides: 31N, 40N, 90N, 120N	ttgacagctagctcagtcctaggtataatactagtATATTAGATATTAGTCATATGACTGACGGANGTGGAGTTNCCACATGAAGTATGACTAGGCATATTATCTTATATGCCACAAAAAGCCGNCCGTCTGGGCAAAAAAAGCCTGGATTGCGNCGGCTTTTTTATatggaaaaggaggaaggatctatgagcaaagga	PAGE Ultramer	J

Supplementary Table 6. Oligonucleotides used for sequencing library preparation.

Below is a table of oligonucleotides used during the cotranscriptional SHAPE-Seq and combinatorial mutagenesis experiments. The 'xxxxxx' sequence in the Illumina primer represents any TruSeq index. For the reverse template amplification primers, the sequence 'NNNNNN…' is the reverse complement from the 3' end of the intermediate length being amplified. Abbreviations within primer sequences are as follows: '/5Biosg/' is a 5' biotin moiety, '/5Phos/' is a 5' monophosphate group, '/3SpC3/' is a 3' 3-carbon spacer group, /3ddC/ is a 3' di-deoxy CTP, VIC and NED are fluorophores (ABI), and asterisks indicate a phosphorothioate backbone modification. These abbreviations were used for compatibility with the Integrated DNA Technologies ordering notation.

Description	Sequence	Purification	ID
RNA linker	/5Phos/CUGACUCGGGCACCAAGGA/3ddC/	None	K
RT primer	/5Biosg/GTCCTTGGTGCCCGAGT	None	L
DNA adapter	/5Phos/AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC/3SpC3/	PAGE	M
QA primer (+)	VIC-GTGACTGGAGTTCAGACGTGTGCTC	None	O
QA primer (-)	NED-GTGACTGGAGTTCAGACGTGTGCTC	None	P
PE_F †	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT	HPLC	Q
Selection primer (+)	CTTTCCCTACACGACGCTCTTCCGATCTRRRYGCATCCACAATAGAAGAAGGATGC*C*G*C*A	None	R
Selection primer (-)	CTTTCCCTACACGACGCTCTTCCGATCTYYYRGCATCCACAATAGAAGAAGGATGC*C*G*C*A	None	S
Illumina reverse primers (TruSeq)†	CAAGCAGAAGACGGCATACGAGATxxxxxxGTGACTGGAGTTCAGACGTGTGCTC	PAGE	T

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