



HHS Public Access

Author manuscript

Nat Chem Biol. Author manuscript; available in PMC 2020 April 21.

Published in final edited form as:

Nat Chem Biol. 2019 November ; 15(11): 1067–1076. doi:10.1038/s41589-019-0382-7.

A ligand gated strand displacement mechanism for ZTP riboswitch transcription control

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

¹Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL 60208, USA

²Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL 60208, USA

³Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, 14853, USA

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

Abstract

Cotranscriptional folding is an obligate step of RNA biogenesis that can guide RNA structure formation and function through transient intermediate folds. This process is particularly important for transcriptional riboswitches in which the formation of ligand-dependent structures during transcription regulates downstream gene expression. However, the intermediate structures that comprise cotranscriptional RNA folding pathways and the mechanisms that enable transit between them remain largely unknown. Here we determine the series of cotranscriptional folds and rearrangements that mediate antitermination by the *Clostridium beijerinckii pfl* ZTP riboswitch in response to the purine biosynthetic intermediate ZMP. We uncover sequence and structural determinants that modulate an internal RNA strand displacement process and identify biases within natural ZTP riboswitch sequences that promote on-pathway folding. Our findings establish a mechanism for *pfl* riboswitch antitermination and suggest general strategies by which nascent RNA molecules navigate cotranscriptional folding pathways.

Introduction

The coupling of transcription and folding is ubiquitous within RNA biogenesis¹. Nascent RNA folding is directed by the 5' to 3' polarity of transcription and the typically slower rate

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

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

Author Contributions

E.J.S. conceived the study and designed the experiments. E.J.S. and J.B.L. conceived the methodology. E.J.S. performed all RNA structure probing and combinatorial mutagenesis experiments. E.J.S., L.C., and K.E.B., performed targeted in vitro transcription experiments. P.D.C. performed molecular cloning. E.J.S. analyzed and interpreted data. E.J.S. wrote the software. E.J.S. and J.B.L. wrote the manuscript with input from L.C., K.E.B., and P.D.C.

Competing Interests

No competing interests declared

of nucleotide addition relative to base pair formation²⁻⁴. Consequently, cotranscriptional RNA folding favors local structures that can pose energetic barriers to the formation of long-range interactions⁵⁻⁷. The tendency of RNA molecules to enter such kinetic traps is thought to be the basis for gene regulation by riboswitches, non-coding RNAs that adopt alternate conformations to control gene expression in response to chemical ligands^{8,9}. The study of riboswitches has consistently revealed diverse roles for RNA molecules in cellular physiology^{9,10}. Furthermore, riboswitches are emerging as antibiotic targets¹¹, diagnostic biosensors^{12,13}, and imaging tools¹⁴.

Riboswitch architecture typically comprises a ligand-sensing aptamer and an 'expression platform' that together direct a regulatory outcome based on ligand occupancy⁸. These domains typically overlap such that mutually exclusive aptamer and expression platform structures block or allow gene expression⁸. Cotranscriptional folding is crucial for riboswitches that regulate transcription because ligand recognition must occur within a limited window before expression platform folding¹⁵. While atomic-resolution structures of diverse riboswitch aptamers have provided a detailed understanding of ligand-aptamer complexes⁹, the series of folding intermediates that mediate riboswitch function has only been described in a handful of cases¹⁶⁻¹⁹. Thus, understanding how ligand-dependent aptamer stabilization controls expression platform folding during transcription remains a major goal.

Here we investigated how a riboswitch that senses the purine biosynthetic intermediate 5-aminoimidazole-4-carboxamide riboside 5'-triphosphate (ZTP) mediates transcription antitermination²⁰. The discovery of the ZTP riboswitch revealed how ZTP and its monophosphate derivative ZMP (Fig. 1a) function as bacterial alarmones for 10-formyl-tetrahydrofolate (10f-THF) deficiency^{20,21}. The ZTP aptamer comprises a helix-junction-helix motif (P1-J1/2-P2) and a small hairpin (P3) that are separated by a variable linker but interact through a pseudoknot (PK) to form the ZTP binding pocket^{20,22-24} (Fig. 1a, b). In the *Clostridium beijerinckii* (*C. beijerinckii*, *Cbe*) *pfl* ZTP riboswitch, the 5' half of the intrinsic terminator stem comprises the entire P3/L3 hairpin and forms a pseudoknot with J1/2²⁰ (Fig. 1a). Crystallographic studies of ligand-bound ZTP aptamers revealed extensive contacts between the aptamer subdomains that could stabilize the aptamer against transcription termination²²⁻²⁴ (Fig. 1b). However, the ZMP-dependence of these interactions during transcription and the precise antitermination mechanism remains unclear.

Our analysis of the *pfl* riboswitch by nascent RNA structure probing revealed ligand-dependent folding pathways with three notable features: 1) a transient intermediate precedes aptamer folding, 2) aptamer folding is ZMP-independent but ZMP binding stabilizes a network of tertiary interactions, and 3) ZMP binds within a narrow transcription window that may be extended by transcription pauses. Functional analysis using combinatorial mutant libraries revealed that ZMP binding antiterminates transcription by controlling intrinsic terminator nucleation and strand displacement through a pseudoknotted aptamer hairpin. Furthermore, analysis of diverse ZTP riboswitch sequences uncovered context-dependent sequence preferences that may avoid off-pathway folds. Our findings reveal the mechanism of *pfl* riboswitch antitermination and suggest general principles that could govern cotranscriptional RNA folding.

Results

Transcription rate tunes the *pfl* riboswitch ZMP response

The *Cbe pfl* ZTP riboswitch was shown to be functional when transcribed *in vitro* by *Escherichia coli* (*E. coli*, *Eco*) RNA polymerase (RNAP) under limiting NTP conditions²⁰. To evaluate *pfl* riboswitch control of *Eco* RNAP transcription when nucleotide addition is not limited, we measured ZMP-mediated antitermination as a function of NTP concentration. Consistent with a kinetic model of riboswitch control¹⁵, slower transcription at 100 μ M NTPs reduced basal terminator readthrough and increased antitermination relative to faster transcription at 500 μ M NTPs (Fig. 1c, Supplementary Fig. 1a). Conversely, inclusion of *Eco* NusA promoted termination²⁵ (Fig. 1d, Supplementary Fig. 1b). In all conditions 1 mM ZMP promoted substantial antitermination over basal terminator readthrough but did not saturate the riboswitch response. Importantly, the *pfl* riboswitch antitermination response occurs primarily from 0.1 to 1 mM ZMP when transcribed by *Eco* RNAP with 500 μ M NTPs or NusA. This range approximates the Z nucleotide pool in *Salmonella typhimurium* and *E. coli* before and after psicofuranine-induced folate stress^{21,26}. Thus, although *Eco* RNAP may differ from *Cbe* RNAP in properties such as speed, pausing, and nascent RNA interactions^{5,27}, our measurements below describe a functional ZTP riboswitch.

Consensus pauses punctuate *pfl* riboswitch transcription

Transcription pausing²⁸ can facilitate riboswitch folding and ligand binding kinetics^{15,17,19} and can be sensitive to aptamer state²⁹. Although pause sequence recognition can be species specific²⁷, diverse multi-subunit RNAPs can recognize a consensus pause sequence (G₋₁₁, G₋₁₀, Y₋₁, G₊₁; -1 corresponds to the RNA 3' end)^{30,31}. We therefore performed single-round *in vitro* transcription with limiting GTP to extend pause lifetime at consensus pause sites^{30,31}. These conditions exposed three consensus pauses that map to positions U96 in L3, and C118 and C121 in the terminator stem (Fig. 1e, f and Supplementary Fig. 1d). We did not observe any pauses that are strictly ZMP- or NusA-dependent (Supplementary Fig. 1c). Notably, at the U96 pause only the P1-J1/2-P2 aptamer subdomain has emerged from RNAP and at the C118 and C121 pauses the entire aptamer has emerged. The observed pauses could therefore extend the time for P1 folding and for pseudoknot folding and ZMP binding, respectively.

A transient structure precedes *pfl* aptamer folding

We previously developed cotranscriptional Selective 2'-Hydroxyl Acylation analyzed by Primer Extension Sequencing (SHAPE-Seq), which couples high-throughput RNA chemical probing^{32,33} with *in vitro* transcription to systematically characterize nascent RNA structures^{18,34}. Our experiment uses a DNA template pool with randomly placed biotin-streptavidin roadblocks to distribute *Eco* RNAP across every template position³⁴. The nascent RNAs displayed by these stalled elongation complexes are then chemically probed to map structures for all intermediate transcripts in a single reaction^{18,34}. To uncover *pfl* riboswitch folding intermediates we performed cotranscriptional SHAPE-Seq with and without 1 mM ZMP (Fig. 2a–d). Initial *pfl* aptamer folding is defined by the ZMP-independent formation of an intermediate hairpin (IH1) from nts 12–25 at transcript length

~45 when reduced reactivity in nts 14–16 produces a ‘low-high-low’ reactivity pattern that is characteristic of a simple hairpin¹⁸ (Fig. 2a, b). IH1 persists through transcript ~70, during which P2 folding is observed as increased reactivity at nts 38–40 flanked by low reactivity at nts 32–35 and 41–44 (Fig. 2a–c). From transcript lengths ~70–81, IH1 loop reactivity decreases as J1/2 (nts 21–31) reactivity increases, suggesting rearrangement of IH1 to form P1 (Fig. 2a, c, e). Given the ~14 nt footprint of *Eco* RNAP on nascent RNA³⁵, this transition correlates with greater favorability for P1 relative to IH1 at transcript length 56 when equilibrium refolded (Supplementary Fig. 2). Because cotranscriptional SHAPE-Seq probes RNAs within roadblocked elongation complexes, IH1 folding may be enabled by transcription arrest. However, because local RNA folding is typically orders of magnitude faster than nucleotide addition by bacterial RNAPs¹, the persistence of IH1 for at least 35 nt addition cycles suggests that it can form during uninterrupted transcription. Minimum free energy structure prediction³⁶ indicated that ~50% of 532 ZTP riboswitches from bacterial genomes²⁰ have the capacity for an intermediate structure as favorable as *Cbe* IH1 (Supplementary Fig. 3). Randomized control predictions suggest the capacity of natural ZTP aptamer sequences for IH1-like structures is a consequence of the high GC content in the J1/2 pseudoknot sequence, but that IH1 is not an encoded motif in its own right (Supplementary Fig. 3).

***pfl* aptamer pseudoknot folding is ZMP-independent**

Following P1 folding, we observed the ZMP-independent formation of a hairpin spanning nts 59–78, indicated by high reactivity from nts 66–71 flanked by regions of low reactivity, and an adjacent 7 nt unstructured region that comprise the linker between P1 and P3 (Fig. 2a, c). ZMP-independent pseudoknot folding is then observed as decreased reactivity at nts 25–29 across transcript lengths 106–112, as L3 emerges from *Eco* RNAP (Fig. 2a, e). In agreement, equilibrated RNA intermediates form the pseudoknot at transcript 95 when complete pairing between nts 25–29 and 91–95 is possible (Supplementary Fig. 2). These data indicate that the pseudoknot can fold before P3, though the order of PK and P3 folding may differ during uninterrupted transcription. Importantly, coordinated reactivity changes at multiple nucleotides from transcript lengths ~117–119 suggests that ZMP can bind only after P3 has emerged from *Eco* RNAP (Fig. 2d–f). In agreement, SHAPE probing of equilibrium refolded RNAs reveals ZMP-dependent reactivity changes when P3 is expected to fold at transcript lengths 99 and 100 (Supplementary Fig. 2). Furthermore, mutations that disrupt and restore pseudoknot base pairs disrupt and restore pseudoknot folding and ZMP-binding, respectively (Supplementary Figs. 4 and 5). The requisite folding of P3 before observation of ZMP-dependent SHAPE reactivity differences does not exclude the possibility that ZMP could bind earlier during transcription. However, our measurements are consistent with a previous finding that the P1 subdomain has no affinity for ZMP unless P3 is supplied in *trans*²³.

ZMP binding stabilizes *pfl* aptamer tertiary interactions

Comparison of nascent ‘apo’ and ‘holo’ intermediate transcripts revealed ZMP-responsive nucleotides that agree with equilibrium in-line probing measurements²⁰ and can be categorized as P1- or pseudoknot-associated by comparing to the ZMP-bound crystal structures of other ZTP aptamers^{22–24} (Fig. 2f and Supplementary Fig. 6). In both

cotranscriptional and equilibrium conditions we observe a coordinated ZMP-dependent reactivity decrease across nts G21, A22, A45 and nts 47–49 beginning at transcript lengths ~117 and 100, respectively (Fig. 2f and Supplementary Fig. 6a, c). These signatures suggest that formation of a primarily non-Watson-Crick (WC) helical extension of P1^{22–24} depends on ZMP binding. Further ZMP-dependent reactivity changes occur simultaneously at nts A30, A31, U39, and A40, which directly contact or are proximal to pseudoknot base pairs^{20,22–24} (Fig. 2f, Supplementary Fig. 6b, d). Decreased reactivity at U39 and A40 is consistent with formation of a Type I A-minor interaction between A40 and the J1/2:P3 pseudoknot^{22–24}, whereas increased reactivity at A30 may be due to formation of a bulge upon stacking between A31 and the G29:C91 pseudoknot base pair²⁴. Pseudoknot disruption renders the above nucleotides ZMP-non-responsive and restoration of pseudoknot base pairs recovers detectable, but weaker, signatures of binding relative to the WT sequence (Supplementary Figs. 4, 5, 6). The ZMP-dependent stabilization of contacts between A40/A31 and the pseudoknot suggests that ZMP-binding coordinates a P2 conformation that promotes formation of non-WC P1 base pairs and the P1:P3 ribose zipper^{20,22–24}. While we observe ZMP recognition as coordinated reactivity changes across many nucleotides, our data cannot distinguish whether these changes happen in concert or in a series of folding events.

ZMP-binding kinetically traps the *pfl* aptamer

We next observed bifurcation of the riboswitch folding pathway into terminated and antiterminated states (Fig. 2e). Without ZMP, the first signature of terminator folding is a gradual increase in reactivity at nts 25–29 in J1/2 from transcript lengths ~125–132, suggesting that pseudoknot disruption can begin as the 3' terminator stem emerges from *Eco* RNAP (Fig. 2a, e). In contrast, sustained low reactivity at J1/2 with ZMP suggests that the pseudoknot remains stable (Fig. 2a, e). The exception to this trend is the primary termination site at nt 132 (Supplementary Fig. 7b), where terminated transcripts increase J1/2 reactivity even with ZMP (Fig. 2e). Notably, the C118 and C121 consensus pauses overlap the ~7 nt ligand binding window and could lengthen the time for ZMP recognition (Fig. 2e).

The final riboswitch fold was obscured by Superscript III reverse transcriptase (SSIII) stalling in transcripts beyond the terminator. We therefore performed a targeted experiment using Superscript IV reverse transcriptase (SSIV) to resolve these transcripts (Fig. 2a). Consistent with the observation that reverse transcriptases can have distinct adduct detection biases³⁷, SSIV was less sensitive to adduct detection by cDNA truncation at some nucleotides but agreed overall with SSIII reactivity trends (Fig. 2a). The reduced reactivity at G25, A31, and A40 observed with ZMP after transcript length 117 is maintained across all post-termination lengths, suggesting that the ZMP-bound ON state persists (Fig. 2a, e). In contrast, terminator readthrough without ZMP yields transcripts with high reactivity at nts G25, A31, A40 suggesting that the pseudoknot is disrupted and ZMP is not bound (Fig. 2a, e). Importantly, equilibrated RNAs default to the terminator fold regardless of ZMP condition, suggesting that the antiterminated fold is only accessible cotranscriptionally (Supplementary Fig. 2).

Combinatorial mutagenesis to perturb RNA folding

Cotranscriptional SHAPE-seq identifies intermediate nascent RNA structures but does not reveal the mechanisms of transit between these structures. To ask how specific nucleotide interactions mediate *pfl* riboswitch folding and antitermination, we implemented a combinatorial mutagenesis strategy that perturbs RNA folding by randomizing defined nucleotide groups in long synthetic oligonucleotides (Supplementary Fig. 7a). By measuring antitermination across these variants, we can infer the sequence determinants that direct folding transitions. Our strategy was inspired by a comprehensive analysis of glycine riboswitch point mutations³⁸ and other *in vitro* transcription approaches that systematically perturb RNA transcripts³⁹, and is similar to a recent in-cell fluorescence-based genetic screen for riboswitch function⁴⁰. We find that sequencing measurements approximate those made by gel electrophoresis and are highly reproducible (Supplementary Fig. 7b–e).

Labile P3 base pairing is critical for termination

Cotranscriptional structure probing suggests that P3 and the pseudoknot fold independently of ZMP and are disrupted during termination. If termination requires P3 to be unfolded, terminator hairpin nucleation could begin by formation of a base pair with the 3' most nucleotide of P3 to close the apical terminator loop (Fig. 1a, Supplementary Fig. 8a). In support of this model, a point mutation (G108C) designed to disrupt terminator nucleation into P3 increases fraction readthrough without ZMP by ~30–40% but could not be rescued because compensatory mutations disrupt the highly conserved P3 stem (Supplementary Fig. 8b–d). We therefore systematically assessed how P3 modulates terminator nucleation using a mutagenesis library that extends P3 by up to two base pairs while preserving terminator base pairing and the ZTP aptamer consensus sequence²⁰ (Fig. 3a). For simplicity, we refer to the first and second extended P3 base pairs as 'Pair 4' and 'Pair 5' to describe their position in P3, and to the corresponding terminator nucleotides as 'Invader 4' and 'Invader 5' to reflect the P3 pair they are expected to compete with during terminator nucleation (Fig. 3a). The resulting variants displayed consistently high (0.68–0.92) terminator readthrough with ZMP, but a range of readthrough without ZMP (0.12–0.86) (Fig. 3b). Classification of P3 variants by Pair 4 revealed that stronger P3 stems tend to increase readthrough without ZMP (Fig. 3c). For each Pair 4 variant, a complementary Invader 4 nucleotide reduced readthrough relative to a mismatch (Fig. 3c). Interestingly, variants in which an Invader 5 G nucleates the terminator hairpin at an unpaired 3' C in Pair 5 always function optimally within each sequence group, but not when the G-C pair orientation is reversed (Fig. 3c, orange points, Supplementary Fig. 8f–h). Perfectly matched invading nucleotides yield highly functional riboswitches except when extending P3 by two G-C pairs made it inaccessible to terminator nucleation (Supplementary Fig. 8e). Weak or mismatched invading pairs reduce termination efficiency in the absence of ZMP and remain sensitive to strong P3 base pairing (Supplementary Fig. 8e). Together, these data suggest that labile P3 base pairing is critical to terminator hairpin nucleation.

Termination requires pseudoknot disruption

If terminator nucleation primarily begins at P3, terminator hairpin folding should require strand displacement through P3 and the pseudoknot. Consequently, changing the efficiency

of strand displacement through the pseudoknot should modulate *pfl* riboswitch function. We therefore designed a mutagenesis library that varied R29 and R26 in the pseudoknot (R=A,G), the corresponding pair positions Y91 and Y94 (Y=U,C) of L3, and R119 and R114 in the terminator such that the aptamer consensus sequence is preserved²⁰ (Fig. 4a). These variants displayed a range of fraction readthrough (1mM – 0mM ZMP) from 0.037 to 0.39 (wild-type = 0.39) (Supplementary Fig. 9a) and produced several expected trends: Terminator efficiency depends on the position and severity of mismatches, pseudoknot mismatches reduce ZMP-responsiveness, and optimal variants have complete or near-complete pseudoknot and terminator base pairs (Fig. 4b,c and Supplementary Fig. 9b–e). We assessed the role of strand displacement in termination by asking how pseudoknot and terminator mismatches change the function of a variant with a five GC pair pseudoknot. When the pseudoknot is fully paired, both poly-U proximal (A119:C91) and poly-U distal (A114:C94) terminator mismatches increased fraction readthrough without ZMP to >0.58 (Fig. 4d). Poly-U proximal mismatches are known to destabilize the terminator hairpin stem^{38,41} and all variants with a poly-U proximal mismatch had high fraction readthrough independent of ZMP (Fig. 4d–g). Conversely, a poly-U distal mismatch may interfere with strand displacement rather than cause terminator dysfunction. In this model, a poly-U distal mismatch would interrupt strand displacement after one pseudoknot base pair is disrupted, but terminator function should be recovered by pseudoknot perturbations that permit strand displacement (Fig. 4h–k). Concordantly, the A26:C94 and A29:C91 pseudoknot mismatches compensated for the poly-U distal terminator mismatch individually to yield functional switches, and in combination to yield a ZMP-nonresponsive switch (Fig. 4e–g). Importantly, each single pseudoknot mutant retained at least partial capacity for ZMP-mediated antitermination, and therefore pseudoknot formation (Fig. 4e, f). The trends described above were observed for two other pseudoknot configurations where interruption of strand displacement by mismatches caused a more severe defect than the presence of G:U wobble pairs in the pseudoknot (Supplementary Fig. 9f–i). Together, these data suggest that the apo *pfl* aptamer can form a stable pseudoknotted fold during transcription that must be broken during termination.

***pfl* aptamer mutants are prone to misfolding**

Having identified strand displacement through the pseudoknot as a step in terminator hairpin folding, we next randomized the pseudoknot-contacting nucleotides A31, A40, and A90, alongside U120 to complement N90 in the terminator stem (Supplementary Fig. 10a, b). Surprisingly, all non-WT 90:120 pairs increased terminator readthrough without ZMP and Y90:R120 pairs functioned comparably to mismatches (Supplementary Fig. 10c). We identified three possible causes for these defects: First, U90, H120 variants may interfere with termination by extending the natural pseudoknot or by forming an alternate pseudoknot (Supplementary Fig. 10c). Supporting this interpretation, N31 and N40 variants that are predicted to occlude the pseudoknot through misfolding partially restore termination efficiency for U90, H120 variants (Supplementary Fig. 10c,d). Second, mutations that rescue U90, H120 variants do not rescue U90, G120 variants, suggesting that G120 terminators are defective, possibly due to shifted base pairs (Supplementary Fig. 10c). Third, C90 is predicted to cause an energetically favorable P3 misfold that could resist terminator nucleation (Supplementary Fig. 10c). In addition, U31 aptamer variants increased

readthrough of R90, H120 and U90, H120 terminators (Supplementary Fig. 10c). One explanation for this defect is that U31 can extend P2 two base pairs which may enable P2 to stack with the pseudoknot and disfavor strand displacement during terminator folding (Supplementary Fig. 10d, rightmost panel). The influence of distal aptamer mutations on terminator function further illustrates the dependence of termination on aptamer folding even in the absence of ZMP.

Context-dependent sequence composition guides P3 folding

Of the terminator variants described above, A90C is notable because its position corresponds to the only P3/L3 nucleotide conserved for its presence but not identity²⁰ and because it may favor a P3 misfold (Fig. 5a, b). Importantly, the predicted A90C P3 misfold precludes pseudoknot folding and should therefore be disfavored among all ZTP aptamers regardless of expression platform (Fig. 5b). To evaluate this hypothesis, we asked whether “*Cbe*-like” ZTP aptamers are biased against a C at L3 nucleotide 2 (L3n2, corresponds to *Cbe pfl* nt 90). Remarkably, *Cbe*-like P3/L3 sequences are enriched for A and depleted for C when P3 is 3 bp (~91% A, 0% C; n=102) or 4 bp (~78% A, ~16% C; n=45), but not when P3 is 5 bp (~22% A, ~66% C; n = 41) (Fig. 5c). This suggests that the sequence constraints that avoid the potential for P3 misfolding may not be necessary if P3 is highly favorable. Conversely, P3/L3 sequences that are not *Cbe*-like were enriched for pyrimidines at L3n2 for all P3 lengths (Fig. 5c). These observations suggest that in some sequence and structure contexts the P3/L3 sequence is under selective pressure not only for ZMP recognition, but also for on-pathway folding.

Discussion

Riboswitch control of intrinsic termination requires that aptamer folding and ligand recognition occur before terminator hairpin folding¹⁵ and these same requirements can be critical for translation control¹⁹. In this regard, the ZTP aptamer must overcome the challenge of folding its ligand binding pocket from subdomains that are separated by a non-conserved linker²⁰. We find that the *pfl* aptamer can bind ZMP ~7 nucleotides before terminator nucleation, but that transcription pauses may extend the time for ZMP recognition (Fig. 6). This limited kinetic window may desensitize the *pfl* riboswitch to enable discrimination between an abundant basal Z nucleotide pool (~100 μ M) and folate stress-induced Z nucleotide accumulation (~1–2 mM)^{21,26}. While the precise bounds of the ligand binding window depend on RNAP-specific factors including transcription speed, pausing²⁷, and RNA folding within the RNA exit channel⁴², given the conservation of bacterial multi-subunit RNAPs⁴³ and consensus pausing^{30,31}, we expect our analyses to approximate the natural *pfl* riboswitch folding pathway.

Notably, the apo *pfl* aptamer sequesters the entire upstream terminator stem in a pseudoknotted hairpin. Consequently, termination is only efficient if terminator folding can efficiently undo P3 and pseudoknot base pairs; appending a single base pair to P3 causes severe termination defects. This agrees with an analysis of the *Clostridium tetani* glycine riboswitch type-1 singlet which concluded that expression platform helices must be close in energy to evoke a meaningful ligand response³⁸.

Given that the *pfl* terminator can unfold a pseudoknotted hairpin, how does ZMP binding block termination? ZMP joins P3 and the pseudoknot in a continuous helical stack by forming a Hoogsteen-edge-Watson-Crick edge pair with a conserved U in L3^{22–24}. Furthermore, ZMP recognition promotes non-canonical P1 base pairs that form a ribose zipper with P3^{22–24} both in equilibrium²⁰ and cotranscriptionally (Fig. 2, Supplementary Fig. 2). Thus, P3 is stabilized by two ZMP-dependent structures. In support of this model, we found that terminator efficiency is exquisitely sensitive to competition between P3 base pairs and terminator nucleation (Fig. 3); an extended P3 stem can be disrupted if terminator nucleation does not compete with P3 base pairs. This is consistent with a mechanism proposed by LeCuyer and Crothers for interconversion of mutually exclusive helices by nucleating base pairs that seed formation of a new helix while unwinding an existing helix^{44,45}. This folding mechanism was also recently reported for the *pbuE* adenine riboswitch⁴⁶.

Together, these observations support a *pfl* riboswitch folding pathway that assumes an ON decision. In the absence of ZMP, the favorability of terminator hairpin folding rejects this structural assumption, whereas ZMP binding commits the riboswitch to the initial ON pathway (Fig. 6). This mode of transcription control was also reported for the *Bacillus cereus crcB* fluoride aptamer^{18,47}. In contrast, dynamic pseudoknot folding and unfolding mediate translation control by the *Faecalibacterium prausnitzii* class III preQ₁ riboswitch⁴⁸ and RNA degradation control by the *GlmS* riboswitch⁴⁹.

pfl aptamer folding involves two proposed characteristics of cotranscriptional RNA folding pathways: temporary helices and the avoidance of competitor helices that would result in dysfunctional structures⁵⁰. The formation of transient folds⁷ is unexpected for transcriptional riboswitches because function requires successful and presumably efficient aptamer folding. Interestingly, the IH1 structure is not explicitly encoded in the *pfl* aptamer consensus sequence but is enriched for by the high GC content of J1/2 and the separation of the sequences that comprise P1 by ~30 bp. Whether IH1 is functionally important remains unclear. However, the simplicity of these characteristics suggests that IH1-like structures may be prevalent in other non-coding RNAs. The ZTP aptamer also exhibits a context-dependent sequence preference that may avoid a misfolded alternative structure to P3 that would prevent ZMP recognition (Fig. 5). Importantly, this sequence constraint depends on both the capacity for misfolding and the energetic favorability of the correct fold.

Overall, the *pfl* riboswitch illustrates how RNA molecules contend with the challenges of cotranscriptional structure formation to control gene expression. While these findings are limited to the system studied here, they support a general principle that RNA sequences are selected both for their functional structure and for pathways to fold that structure³.

Online Methods

DNA template preparation

Linear DNA templates were prepared by PCR amplification as described⁵¹. 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 1), 2.5 μl of 10 μM oligonucleotide B, D, or E (reverse primer; Supplementary Table 1), 1 μl of Vent Exo-DNA polymerase (New England Biolabs), and 0.5 μl of plasmid DNA (Supplementary Table 2) were subjected to 30 PCR cycles. Randomly biotinylated DNA templates were prepared as above except that each dNTP and corresponding biotin-11-dNTP (Biotium, Fremont, CA; Perkin Elmer, Waltham, MA) were added to a total of 100 nmol such that ~ 1 biotin-11-dNTP is incorporated within the transcribed region of each DNA template³⁴. DNA templates for combinatorial mutagenesis were prepared as above using oligonucleotides G and E (Supplementary Table 1) from gel-purified 'Ultramers' oligonucleotides (Integrated DNA technologies, Coralville, IA) after conversion to double-stranded DNA by 8 PCR cycles 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 1), 0.25 μl of 100 μM oligo E (reverse primer; Supplementary Table 1), 1 μl of 1 μM ultramer oligonucleotide (Supplementary Table 3), and 2 μl of Vent Exo-DNA polymerase and subsequent QIAquick PCR purification (Qiagen, Hilden, Germany). For all DNA template preparations, 100 μl 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 -80°C for 15 min, centrifuged, washed with 1.5 mL 70% EtOH (v/v), dried using a SpeedVac, and dissolved in 30 μl , run on a 1% agarose gel, and extracted using a QIAquick Gel Extraction kit (Qiagen). DNA concentration was determined by a Qubit 3.0 Fluorometer (Life Technologies, Carlsbad, CA).

Radiolabeled *in vitro* transcription

All radiolabeled single-round transcription reactions contained 10 nM DNA template and 0.016 U/ μl *E. coli* RNA polymerase 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)), and 0.2 mg/ml bovine serum albumin (BSA)⁵¹. When present, NusA (provided by Jeffrey Roberts, Cornell University) was included at 500 nM. 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 variable concentration with DMSO concentration fixed at 2% (v/v) in the final reaction. All reactions were 25 μL . Several protocols for *in vitro* transcription were performed: For dose-response curves with and without NusA, open promoter complexes were formed by incubating reactions containing 200 μM High Purity ATP, GTP, CTP, 50 μM UTP (GE Life Sciences, Chicago, IL) and 0.2 $\mu\text{Ci}/\mu\text{L}$ [α -³²P]UTP (Perkin-Elmer) at 37C for 10 min and initiated by adding magnesium chloride (MgCl_2) to 10 mM and rifampicin (Gold Biotechnology, St. Louis, MO) to 10 $\mu\text{g}/\text{mL}$ ⁵¹; reactions proceeded for 5 min before addition of 125 μL of Stop Solution (0.6 M Tris, pH 8.0, 12 mM EDTA). For dose-response curves at variable NTP concentrations, elongation complexes were stalled at +15²⁰ by incubating reactions containing 2.5 μM ATP and GTP, 1.5 μM UTP, 0.2 $\mu\text{Ci}/\mu\text{L}$ [α -³²P]UTP, and 10 mM MgCl_2 at 37C for 10 min before aliquoting to separate tubes containing 5x the desired concentration of NTPs, ZMP, and rifampicin; reactions were incubated at 37C for 5 min before addition of 125 μL of Stop Solution. Synchronization for time-resolved single-round transcription⁵² was performed either by stalling complexes at +15 and chasing with ATP, UTP, and CTP to 100 μM , GTP to 10 μM ²⁸ and rifampicin to 10

positions; under the described purification protocol the ZMP included during initial RNA synthesis should be completely depleted during the two subsequent phased extractions and precipitations, as is evidenced by comparison of the equilibrium refolded and cotranscriptionally-folded matrices (compare Fig. 2a and Supplementary Fig. 2a). After dissolving purified RNAs in 25 μ l of water, equilibrium refolding was performed by denaturing at 95C for 2 min, snap cooling on ice for 1 min, and adding transcription buffer, 500 μ M NTPs, 10 mM MgCl₂, 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 μ l water and 150 μ l TRIzol, extraction according to the manufacturer's protocol and resuspension in 10 μ l of 10% DMSO.

Sequencing library preparation

Sequencing libraries for cotranscriptional SHAPE-seq were prepared either as previously described⁵⁴ 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 protocol.

RNA 3' linker adenylation and ligation

5'-Phosphorylated linker (Oligonucleotide K, Supplementary Table 4) was adenylated using a 5' DNA Adenylation Kit (New England Biolabs) at 20x scale and purified by TRIzol extraction as described⁵⁴. RNA 3' ligation was performed by combining 10 μ l extracted RNAs in 10% DMSO with 0.5 μ l of SuperaseIN (Life Technologies), 6 μ l 50% PEG 8000, 2 μ l of 10X T4 RNA Ligase Buffer (New England Biolabs), 1 μ l of 2 μ M 5'-adenylated RNA linker and mixing by pipetting. 0.5 μ l of T4 RNA ligase 2, truncated KQ (New England Biolabs) was then added and the reaction was mixed again and incubated at 25C for 3 hrs.

Reverse Transcription

Following linker ligation, RNAs were precipitated by adding 130 μ l RNase-free water, 15 μ l 3M NaOAc pH 5.5, 1 μ l 20 mg/ml glycogen, and 450 μ l of 100% EtOH and storing at -80C for 30 min, centrifuged, washed once with 500 μ l 70% EtOH (v/v), and residual ethanol was removed. For Superscript III reverse transcription, samples were resuspended in 10 μ l RNase-free water and 3 μ l of 0.5 μ M reverse transcription primer (Oligonucleotide L, Supplementary Table 4), denatured at 95C for 2 min, incubated at 65C for 5 min, briefly centrifuged, and placed on ice. 7 μ l of Superscript III reverse transcription master mix (containing 4 μ l of 5x First Strand Buffer (Life Technologies), 1 μ l of 100 mM DTT, 1 μ l 10 mM dNTPs, 0.5 μ l RNase-Free Water, and 0.5 μ l Superscript III) was added and mixed before placing each sample at 45C and incubating at 45C for 1 min, 52C for 25 min, and 65C for 5 min. For Superscript IV reverse transcription, RNAs were precipitated as above, resuspended in 9.5 μ l of RNase-free water, and 3 μ l of reverse transcription primer (Oligonucleotide L, Supplementary Table 4), denatured at 95C for 2 min, incubated at 65C for 5 min, briefly centrifuged, and placed on ice. 7.5 μ l of Superscript IV reverse transcription master mix (containing 4 μ l of 5x SSIV Buffer (Life Technologies), 1 μ l of 100 mM DTT, 1 μ l 10 mM dNTPs, 0.5 μ l RNase OUT (Invitrogen, Waltham, MA), and 1 μ l Superscript IV) was added and mixed before being placing each sample at 45C and

incubating at 45C for 1 min, 52C for 25 min, 65C for 5 min, and 80C for 10 min. 1 µl of 4M sodium hydroxide (NaOH) was added and samples were heated at 95C for 5 min to hydrolyze RNA, partially neutralized by 2 µl of 1M hydrochloric acid (HCl), and precipitated by adding 69 µl 100% EtOH and storing at –80C for 15 min, centrifugation for 15 min at 4C, and washing with 500 µl of 70% EtOH. After removing residual ethanol pellets were dissolved in 22.5 µl of RNase-free water.

Adapter ligation

Adapter ligation was performed as described⁵⁴. Briefly, dissolved cDNA was mixed with 3 µl of 10x CircLigase Buffer (Epicentre), 1.5 µl of 50 mM MnCl₂, 1.5 µl of 1 mM ATP 0.5 µl of 100 µM DNA adapter (Oligonucleotide M, Supplementary Table 4), and 1 µl of CircLigase I (Epicentre, Madison, WI), incubated at 60C for 2 h and 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. 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

Sequencing library quality analysis was performed as previously described⁵⁴ by generating fluorescently labeled dsDNA libraries using oligonucleotides O, P, Q and R or S (Supplementary Table 4). Samples were analyzed by capillary electrophoresis using an ABI 3730xl DNA Analyzer.

Preparation of dsDNA libraries for sequencing

Sequencing libraries for cotranscriptional SHAPE-Seq were prepared as described⁵⁴. 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 4), 2 µl of 0.1 µM of channel-specific selection primer (Oligonucleotides R and S, Supplementary Table 4), 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 100 µM primer PE_F (Oligonucleotide Q, Supplementary Table 4) 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, eluted in 20 µl of 1X TE buffer, and quantified using a Qubit 3.0 Fluorometer (Life Technologies). Molarity was estimated using the length distribution observed in Quality Analysis. Sequencing libraries for combinatorial mutagenesis was performed as 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 2×36 or 2×37 bp paired end reads with 30% PhiX. Cotranscriptional SHAPE-Seq data analysis was performed using Spats v1.0.1 as previously described¹⁸ except that one mismatch was permitted during alignment following the observation that truncations at several nucleotides were enriched for a terminal mutation. For Superscript IV data we frequently observed a non-templated T at full length cDNA 3' ends and treated these reads as full length.

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/mm² 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. Reads were aligned 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 12 and 130 of the riboswitch. Alignment to the riboswitch leader (nts 1–11) was not required because of elevated Superscript IV dropoff in this region and inclusion of these reads did not impact measurements (Supplementary Fig. 7f, g). Control analyses that omitted the poly-uridine tract from alignment did not impact measurements (Supplementary Fig. 7f, g), therefore insertion mutants were aligned through two poly-uridine tract nucleotides to permit usage of 150-cycle v3 MiSeq Reagent Kits (Illumina). Single mismatches were permitted in Read 1 beyond nt 130 if the unambiguous mapping as terminated or full length was possible. 3' ends from 130 to 134 were considered terminated and 3' ends ≥ 135 were considered full length. Fraction readthrough was calculated for each variant by dividing the full length read count by the sum of terminated and full length reads.

Urea PAGE and gel image quantification

Radiolabeled RNAs were fractionated using 9%, 10%, or 12% Urea PAGE sequencing gels prepared using the UreaGel System (National Diagnostics, Atlanta, GA). Reactive nucleotides were detected by an Amersham Biosciences Typhoon 9400 Variable Mode Imager and quantified using ImageQuant (GE Life Sciences). For experiments with uniform labeling, band intensity was normalized for incorporation of [α -³²P]UTP by dividing by U nucleotide count for each transcript. To normalize band intensity in experiments with staged transcription, the fraction [α -³²P]UTP/([α -³²P]UTP + UTP) was determined for the walk (P_{walk}) and chase with 100 μM or 500 μM NTPs (P_{chase}) reaction phases and used to calculate labeling efficiency for each band by the equation ($(P_{\text{walk}} * \text{U nucleotides in walk}) + (P_{\text{chase}} * \text{U nucleotides in chase})$); band intensity was divided by labeling efficiency. Labeling is virtually uniform because the probability of [α -³²P]UTP incorporation is high during the walk and low during the chase (~4.25% vs. 0.067% and 0.013%) and full length transcripts contain only four more U nucleotides than terminated transcripts. Fraction readthrough was determined by dividing the normalized terminated band intensity by the sum of the normalized full length and terminated band intensity. Dose-response curves were

fit using GraphPad Prism (GraphPad Software, San Diego, CA). In Figure 1e, pause sites were determined by a 3'-dNTP sequencing ladder; because C121 migrated closely with G122 transcripts, we inferred C121 pause identity both from the sequencing ladder and because the pause is observed as doublet; under the limiting GTP conditions the doublet bands most likely correspond C121 and G122 which both precede a G.

Non-radiolabeled RNAs were resuspended in transcription loading dye, fractionated by Urea PAGE (10 or 12% polyacrylamide), stained with SYBR Gold (Life Technologies), imaged with a Bio-Rad ChemiDoc Touch Imaging System, and quantified with Image Lab (Bio-Rad, Hercules, CA). Bands were normalized for RNA length dividing band intensity by expected product length. Fraction readthrough was determined as above.

RNA structure prediction

RNA structure prediction was performed using the RNAstructure v6.1³⁶ Fold command with default settings. For IH1 (Supplementary Fig. 3), WT sequences from previously identified ZTP riboswitches²⁰ were obtained from the RefSeq⁵⁵ entries used for their original identification and aligned using INFERNAL v1.1.2⁵⁶. The aptamer segment used for structure prediction contains sequence starting two nucleotides upstream of the longest P1 helices in ZTP riboswitch family multiple sequence alignment²⁰ through the last unstructured nt within J1/2 and therefore may exclude leader sequences which could influence transient structure formation. The 3' terminus of this segment was determined using the RNAstructure³⁶ fold command to predict whether sequence between the last pseudoknot base pair and the first P2 base pair could form a structure. In cases where no structure was predicted, the entire sequence between the pseudoknot and P2 was included in the sequence used for IH1 structure prediction. 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 in the ZTP riboswitch multiple sequence alignment. Shuffled randomization was performed by randomly reordering the nucleotides of natural sequences. All randomized data sets match the WT length distribution.

L3n2 Nucleotide Frequency Analysis

P3 hairpin sequences from without insertions were binned by a match to the motif 'NNNNNGNCCNNNNGGCNN' and then binned by the number of predicted contiguous P3 base pairs. Nucleotide frequency at L3 nucleotide 2 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 5. SHAPE-Seq Reactivity Spectra generated in this work have been deposited in the RNA Mapping Database (RMDb)⁵⁷ (<http://rmdb.stanford.edu/repository/>) 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 6. Source data for all figures are available in the Northwestern University Arch Institutional Repository (<https://doi.org/10.21985/N2220T>). Uncropped gel images are shown in Supplementary Figure 12. All other data that support the findings of this paper are available from the corresponding authors upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Robert Batey, Chad Torgerson, Scott Strobel, Christopher Jones, and Irina Artsimovitch for thoughtful discussions; Jeffrey Roberts (Cornell University) for providing *E. coli* NusA protein; John Lis for use of facilities; 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.

References

1. Pan T & Sosnick T RNA folding during transcription. *Annu Rev Biophys Biomol Struct* 35, 161–175, doi:10.1146/annurev.biophys.35.040405.102053 (2006). [PubMed: 16689632]
2. Zhang J & Landick R A Two-Way Street: Regulatory Interplay between RNA Polymerase and Nascent RNA Structure. *Trends Biochem Sci* 41, 293–310, doi:10.1016/j.tibs.2015.12.009 (2016). [PubMed: 26822487]
3. Lai D, Proctor JR & Meyer IM On the importance of cotranscriptional RNA structure formation. *RNA* 19, 1461–1473, doi:10.1261/rna.037390.112 (2013). [PubMed: 24131802]
4. Al-Hashimi HM & Walter NG RNA dynamics: it is about time. *Curr Opin Struct Biol* 18, 321–329, doi:10.1016/j.sbi.2008.04.004 (2008). [PubMed: 18547802]
5. Pan T, Artsimovitch I, Fang XW, Landick R & Sosnick TR Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc Natl Acad Sci U S A* 96, 9545–9550 (1999). [PubMed: 10449729]
6. Heilman-Miller SL & Woodson SA Effect of transcription on folding of the Tetrahymena ribozyme. *RNA* 9, 722–733 (2003). [PubMed: 12756330]
7. Wong TN, Sosnick TR & Pan T Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc Natl Acad Sci U S A* 104, 17995–18000, doi:10.1073/pnas.0705038104 (2007). [PubMed: 17986617]
8. Garst AD, Edwards AL & Batey RT Riboswitches: structures and mechanisms. *Cold Spring Harb Perspect Biol* 3, doi:10.1101/cshperspect.a003533 (2011).

9. McCown PJ, Corbino KA, Stav S, Sherlock ME & Breaker RR Riboswitch diversity and distribution. *RNA* 23, 995–1011, doi:10.1261/rna.061234.117 (2017). [PubMed: 28396576]
10. Nelson JW & Breaker RR The lost language of the RNA World. *Sci Signal* 10, doi:10.1126/scisignal.aam8812 (2017).
11. Howe JA et al. Selective small-molecule inhibition of an RNA structural element. *Nature* 526, 672–677, doi:10.1038/nature15542 (2015). [PubMed: 26416753]
12. Kellenberger CA, Wilson SC, Sales-Lee J & Hammond MC RNA-based fluorescent biosensors for live cell imaging of second messengers cyclic di-GMP and cyclic AMP-GMP. *J Am Chem Soc* 135, 4906–4909, doi:10.1021/ja311960g (2013). [PubMed: 23488798]
13. Porter EB, Polaski JT, Morck MM & Batey RT Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat Chem Biol* 13, 295–301, doi:10.1038/nchembio.2278 (2017). [PubMed: 28092358]
14. Braselmann E et al. A multicolor riboswitch-based platform for imaging of RNA in live mammalian cells. *Nat Chem Biol* 14, 964–971, doi:10.1038/s41589-018-0103-7 (2018). [PubMed: 30061719]
15. Wickiser JK, Winkler WC, Breaker RR & Crothers DM The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* 18, 49–60, doi:10.1016/j.molcel.2005.02.032 (2005). [PubMed: 15808508]
16. Frieda KL & Block SM Direct observation of cotranscriptional folding in an adenine riboswitch. *Science* 338, 397–400, doi:10.1126/science.1225722 (2012). [PubMed: 23087247]
17. Perdrizet GA 2nd, Artsimovitch I, Furman R, Sosnick TR & Pan T Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc Natl Acad Sci U S A* 109, 3323–3328, doi:10.1073/pnas.1113086109 (2012). [PubMed: 22331895]
18. Watters KE, Strobel EJ, Yu AM, Lis JT & Lucks JB Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat Struct Mol Biol* 23, 1124–1131, doi:10.1038/nsmb.3316 (2016). [PubMed: 27798597]
19. Chauvier A et al. Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nat Commun* 8, 13892, doi:10.1038/ncomms13892 (2017). [PubMed: 28071751]
20. Kim PB, Nelson JW & Breaker RR An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol Cell* 57, 317–328, doi:10.1016/j.molcel.2015.01.001 (2015). [PubMed: 25616067]
21. Bochner BR & Ames BN ZTP (5-amino 4-imidazole carboxamide riboside 5'-triphosphate): a proposed alarmone for 10-formyl-tetrahydrofolate deficiency. *Cell* 29, 929–937 (1982). [PubMed: 6185232]
22. Trausch JJ, Marcano-Velazquez JG, Matyjasik MM & Batey RT Metal Ion-Mediated Nucleobase Recognition by the ZTP Riboswitch. *Chem Biol* 22, 829–837, doi:10.1016/j.chembiol.2015.06.007 (2015). [PubMed: 26144884]
23. Jones CP & Ferre-D'Amare AR Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains. *Nat Struct Mol Biol* 22, 679–685, doi:10.1038/nsmb.3073 (2015). [PubMed: 26280533]
24. Ren A, Rajashankar KR & Patel DJ Global RNA Fold and Molecular Recognition for a pfl Riboswitch Bound to ZMP, a Master Regulator of One-Carbon Metabolism. *Structure* 23, 1375–1381, doi:10.1016/j.str.2015.05.016 (2015). [PubMed: 26118534]
25. Schmidt MC & Chamberlin MJ nusA protein of *Escherichia coli* is an efficient transcription termination factor for certain terminator sites. *J Mol Biol* 195, 809–818 (1987). [PubMed: 2821282]
26. Rohlman CE & Matthews RG Role of purine biosynthetic intermediates in response to folate stress in *Escherichia coli*. *J Bacteriol* 172, 7200–7210 (1990). [PubMed: 2254281]
27. Artsimovitch I, Svetlov V, Anthony L, Burgess RR & Landick R RNA polymerases from *Bacillus subtilis* and *Escherichia coli* differ in recognition of regulatory signals in vitro. *J Bacteriol* 182, 6027–6035 (2000). [PubMed: 11029421]

28. Artsimovitch I & Landick R Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc Natl Acad Sci U S A* 97, 7090–7095, doi:10.1073/pnas.97.13.7090 (2000). [PubMed: 10860976]
29. Widom JR et al. Ligand Modulates Cross-Coupling between Riboswitch Folding and Transcriptional Pausing. *Mol Cell* 72, 541–552 e546, doi:10.1016/j.molcel.2018.08.046 (2018). [PubMed: 30388413]
30. Larson MH et al. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* 344, 1042–1047, doi:10.1126/science.1251871 (2014). [PubMed: 24789973]
31. Vvedenskaya IO et al. Interactions between RNA polymerase and the “core recognition element” counteract pausing. *Science* 344, 1285–1289, doi:10.1126/science.1253458 (2014). [PubMed: 24926020]
32. Strobel EJ, Yu AM & Lucks JB High-throughput determination of RNA structures. *Nat Rev Genet* 19, 615–634, doi:10.1038/s41576-018-0034-x (2018). [PubMed: 30054568]
33. Bevilacqua PC & Assmann SM Technique Development for Probing RNA Structure In Vivo and Genome-Wide. *Cold Spring Harb Perspect Biol* 10, doi:10.1101/cshperspect.a032250 (2018).
34. Strobel EJ, Watters KE, Nedialkov Y, Artsimovitch I & Lucks JB Distributed biotin-streptavidin transcription roadblocks for mapping cotranscriptional RNA folding. *Nucleic Acids Res* 45, e109, doi:10.1093/nar/gkx233 (2017). [PubMed: 28398514]
35. Komissarova N & Kashlev M Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc Natl Acad Sci U S A* 95, 14699–14704 (1998). [PubMed: 9843952]
36. Reuter JS & Mathews DH RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* 11, 129, doi:10.1186/1471-2105-11-129 (2010). [PubMed: 20230624]
37. Sexton AN, Wang PY, Rutenberg-Schoenberg M & Simon MD Interpreting Reverse Transcriptase Termination and Mutation Events for Greater Insight into the Chemical Probing of RNA. *Biochemistry* 56, 4713–4721, doi:10.1021/acs.biochem.7b00323 (2017). [PubMed: 28820243]
38. Torgerson CD, Hiller DA, Stav S & Strobel SA Gene regulation by a glycine riboswitch singlet uses a finely tuned energetic landscape for helical switching. *RNA* 24, 1813–1827, doi:10.1261/rna.067884.118 (2018). [PubMed: 30237163]
39. Vvedenskaya IO et al. Massively Systematic Transcript End Readout, “MASTER”: Transcription Start Site Selection, Transcriptional Slippage, and Transcript Yields. *Mol Cell* 60, 953–965, doi:10.1016/j.molcel.2015.10.029 (2015). [PubMed: 26626484]
40. Polaski JT, Kletzien OA, Drogalis LK & Batey RT A functional genetic screen reveals sequence preferences within a key tertiary interaction in cobalamin riboswitches required for ligand selectivity. *Nucleic Acids Res* 46, 9094–9105, doi:10.1093/nar/gky539 (2018). [PubMed: 29945209]
41. Larson MH, Greenleaf WJ, Landick R & Block SM Applied force reveals mechanistic and energetic details of transcription termination. *Cell* 132, 971–982, doi:10.1016/j.cell.2008.01.027 (2008). [PubMed: 18358810]
42. Hein PP et al. RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat Struct Mol Biol* 21, 794–802, doi:10.1038/nsmb.2867 (2014). [PubMed: 25108353]
43. Geszvain KM & Landick R in *American Society for Microbiology* (ed Higgins NP) 283–396 (Washington, DC., 2004).
44. LeCuyer KA & Crothers DM Kinetics of an RNA conformational switch. *Proc Natl Acad Sci U S A* 91, 3373–3377, doi:10.1073/pnas.91.8.3373 (1994). [PubMed: 8159754]
45. LeCuyer KA & Crothers DM The *Leptomonas collosoma* spliced leader RNA can switch between two alternate structural forms. *Biochemistry* 32, 5301–5311 (1993). [PubMed: 8499434]
46. Drogalis LK & Batey RT Requirements for efficient cotranscriptional regulatory switching in designed variants of the *Bacillus subtilis* pbuE adenine-responsive riboswitch. *bioRxiv*, doi:10.1101/372573 (2018).
47. Zhao B, Guffy SL, Williams B & Zhang Q An excited state underlies gene regulation of a transcriptional riboswitch. *Nat Chem Biol* 13, 968–974, doi:10.1038/nchembio.2427 (2017). [PubMed: 28719589]

48. Liberman JA et al. 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, E3485–3494, doi:10.1073/pnas.1503955112 (2015). [PubMed: 26106162]
49. Savinov A & Block SM Self-cleavage of the glmS ribozyme core is controlled by a fragile folding element. *Proc Natl Acad Sci U S A* 115, 11976–11981, doi:10.1073/pnas.1812122115 (2018). [PubMed: 30397151]
50. Meyer IM & Miklos I Co-transcriptional folding is encoded within RNA genes. *BMC Mol Biol* 5, 10, doi:10.1186/1471-2199-5-10 (2004). [PubMed: 15298702]

Methods-only References

51. Strobel EJ & Roberts JW Two transcription pause elements underlie a sigma70-dependent pause cycle. *Proc Natl Acad Sci U S A* 112, E4374–4380, doi:10.1073/pnas.1512986112 (2015). [PubMed: 26216999]
52. Landick R, Wang D & Chan CL Quantitative analysis of transcriptional pausing by *Escherichia coli* RNA polymerase: his leader pause site as paradigm. *Methods Enzymol* 274, 334–353 (1996). [PubMed: 8902817]
53. Mortimer SA & Weeks KM A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J Am Chem Soc* 129, 4144–4145, doi:10.1021/ja0704028 (2007). [PubMed: 17367143]
54. Watters KE, Yu AM, Strobel EJ, Settle AH & Lucks JB 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 (2016). [PubMed: 27064082]
55. Pruitt KD, Tatusova T & Maglott DR NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res* 35, D61–65, doi:10.1093/nar/gkl842 (2007). [PubMed: 17130148]
56. Nawrocki EP & Eddy SR Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* 29, 2933–2935, doi:10.1093/bioinformatics/btt509 (2013). [PubMed: 24008419]
57. Cordero P, Lucks JB & Das R An RNA Mapping DataBase for curating RNA structure mapping experiments. *Bioinformatics* 28, 3006–3008, doi:10.1093/bioinformatics/bts554 (2012). [PubMed: 22976082]

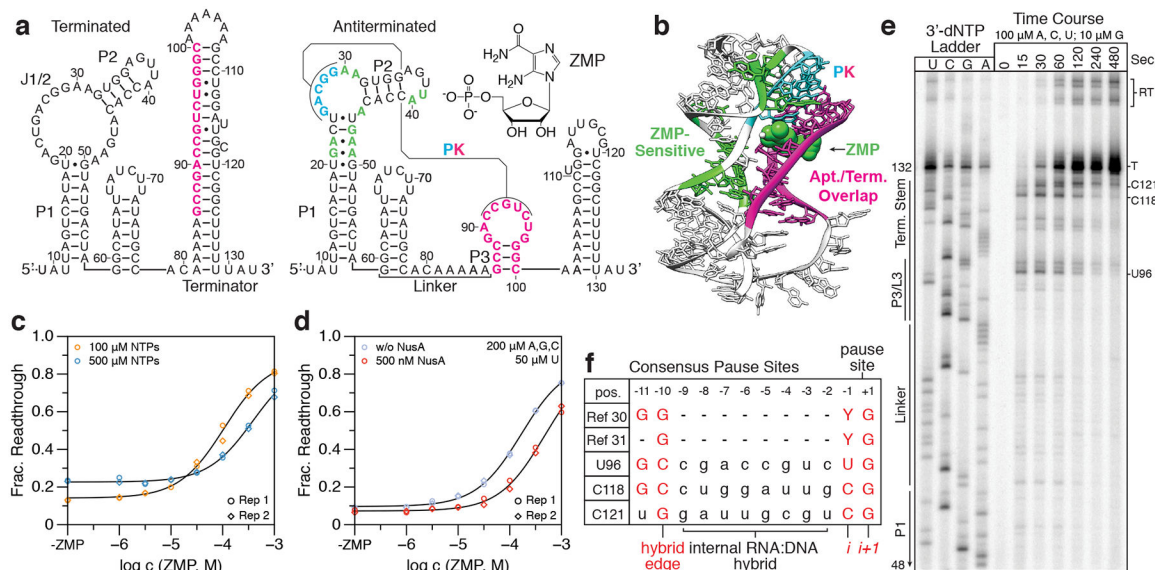


Figure 1. Overview and *in vitro* characterization of the *C. beijerinckii pfl* riboswitch.

(a) Secondary structure of the *Cbe pfl* riboswitch terminated and antiterminated folds^{20,22–24}. Aptamer/terminator overlap is magenta, J1/2 pseudoknot nucleotides are cyan, and ZMP-responsive nucleotides (Figure 2) are green.

(b) Crystal structure of the *Thermosinus carboxydivorans pfl* riboswitch (PDB: 4ZNP)²⁴ highlighted to match corresponding nucleotide regions in (a).

(c-d) ZMP dose-response curves for the *Cbe pfl* riboswitch measured by single-round *in vitro* transcription with 100 μ M and 500 μ M NTPs (c) and \pm 500 nM NusA (d).

Representative gels are shown in Supplementary Figs. 1a and 1b.

(e) Time-resolved single-round *in vitro* transcription of the *Cbe pfl* riboswitch without ZMP.

(f) Sequences of consensus pause sites^{30,31} identified in (e).

Panels (c) and (d) are from $n=2$ independent replicates; for panel (d) 316 μ M and 1 mM ZMP data were obtained separately from values for 100 μ M ZMP. Panel e is $n=1$ and agrees with comparable measurements in the same and additional synchronization, ZMP, and NusA conditions in Supplementary Fig. 1c, d. Uncropped source gels are shown in Supplementary Fig. 12.

Source data are available in the Northwestern University Arch Institutional Repository (<https://doi.org/10.21985/N2220T>).

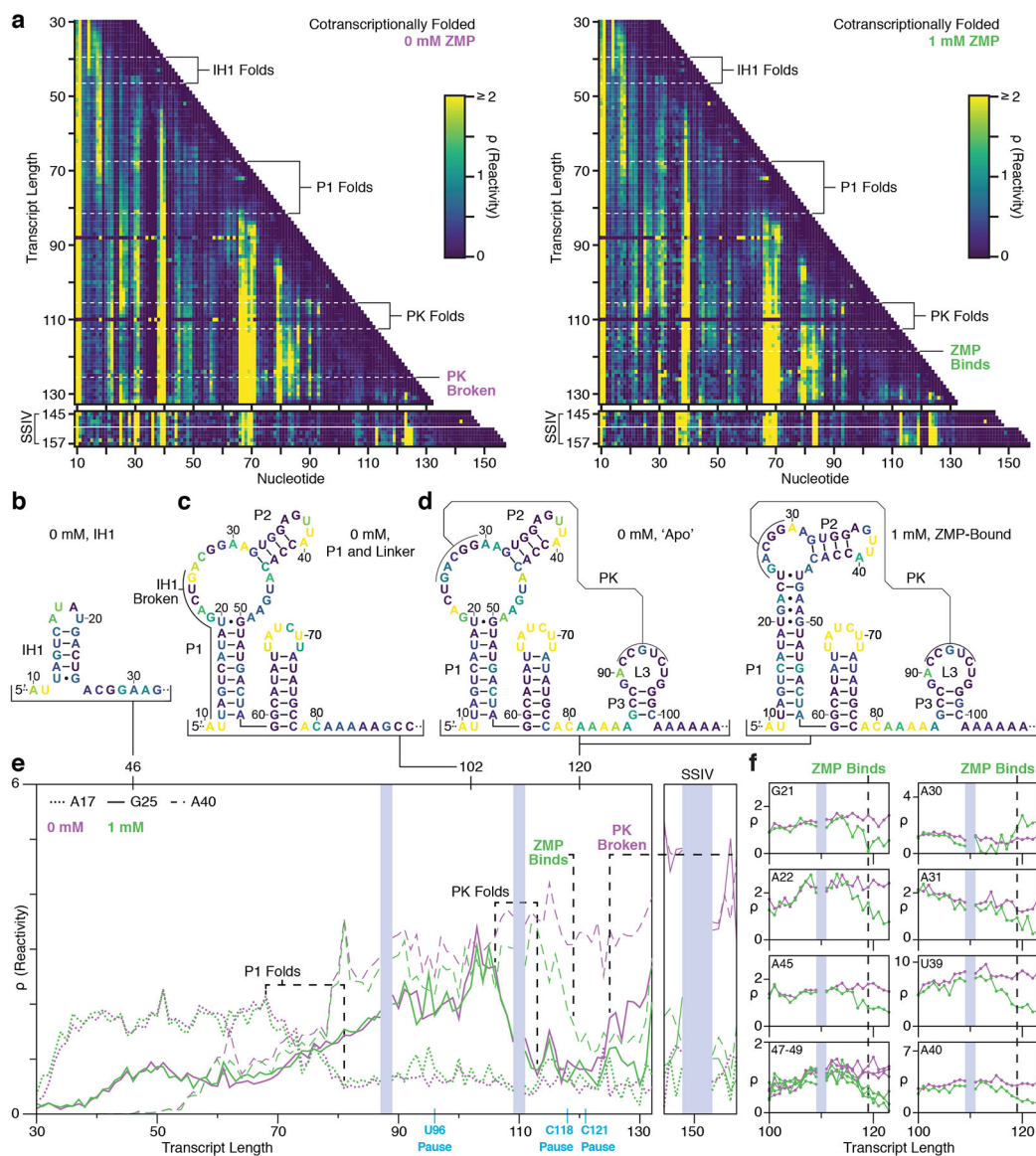


Figure 2. *C. beijerinckii pfl* riboswitch folding intermediates

(a) Cotranscriptional SHAPE-Seq reactivity matrices for the *Cbe pfl* riboswitch with 0 mM and 1 mM ZMP. The unstructured leader (nts 1–9) is not shown. Data for transcripts 88 and 110 are absent due to ambiguous 3' end mapping of sequencing reads. Reactivity spectra for transcripts 145–148 and 153–157 are from a separate targeted experiment using Superscript IV reverse transcriptase. Reactivity (ρ) is capped at 2 for heatmap presentation.

(b-d) Intermediate hairpin (IH1) (b), P1-J1/2-P2 and linker (c), and apo and ZMP-bound aptamer (d) secondary structures colored by reactivity from transcripts 46 (b), 102 (c), and 120 (d). Structures were inferred from manual reactivity analysis and MFE prediction informed by covariation²⁰ and crystal structures^{22–24}. Nucleotides within the RNAP footprint are omitted.

(e-f) Transcript length-dependent reactivity changes for select nucleotides showing folding transitions (e) and ZMP-responsive nucleotides (non-canonical P1 pairs, left; pseudoknot-contacting, right) (f). Transcripts 88 and 110 are omitted as described in (a).

All data in panels b-f are from (a). Upper and lower matrices in (a) are representative of $n=3$ and $n=2$ independent replicates, respectively. Replicate data and correlations are in Supplementary Fig. 11.

Source data are available in the Northwestern University Arch Institutional Repository (<https://doi.org/10.21985/N2220T>).

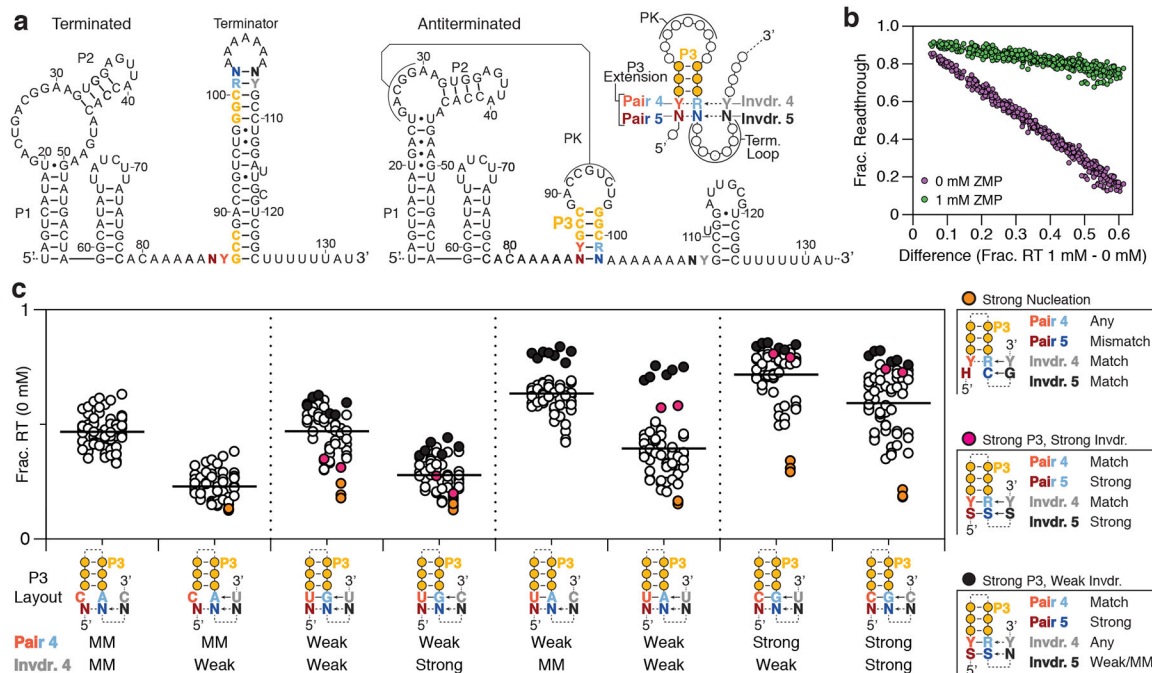


Figure 3. Mutagenesis of the *C. beijerinckii* pfl aptamer P3 stem.

(a) *Cbe pfl* riboswitch terminated and antiterminated secondary structures depicting P3 randomization scheme; randomized nucleotides are color coded. Insertion mutations are not counted in nucleotide numbering for consistency. The proposed P3 stem/terminator base pairing competition is shown. Potential P3 and invading base pairs are dashed; invading nucleotides are labeled 'Invdr'.

(b) Fraction readthrough for P3 variants with 0 mM and 1 mM ZMP ordered by difference (1mM - 0mM).

(c) Fraction readthrough with 0mM ZMP grouped by Pair 4 and Invader 4 identity. Horizontal lines are group averages. P3 Layout is colored as in (b). Notable Pair 4/Invader 4 configurations are depicted to the right; variable nucleotides in these configurations 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). Required base pairs are solid and optional base pairs are dashed. All data in panels (b) and (c) are the average of $n=2$ independent replicates of each ZMP condition for the mutagenesis library depicted in (a). Replicates are compared in Supplementary Figure 7c.

Source data are available in the Northwestern University Arch Institutional Repository (<https://doi.org/10.21985/N2220T>).

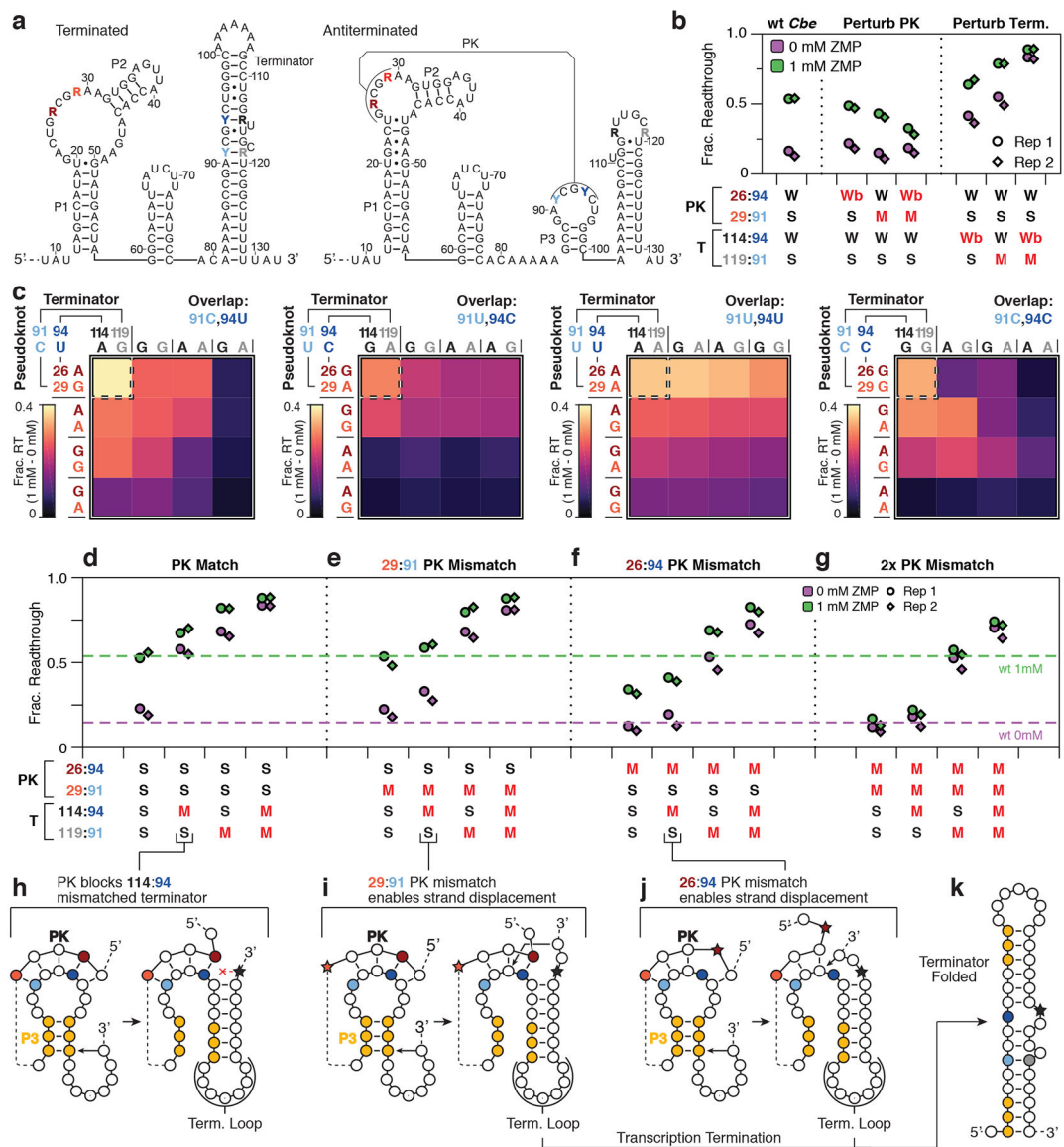


Figure 4. Mutagenesis of *C. beijerinckii pfl* aptamer pseudoknot and terminator base pairs

(a) *Cbe pfl* riboswitch terminated and antiterminated secondary structures depicting pseudoknot randomization scheme; randomized nucleotides are color coded. R=A,G; Y=C,U.

(b) Fraction readthrough for select pseudoknot mutants with 0 mM or 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 indicates deviation from the wild-type pairs.

(c) Heat maps showing the difference in fraction readthrough (1 mM – 0 mM) for pseudoknot mutants grouped by aptamer/terminator overlap sequence (nts 91 and 94). Variants with perfect pseudoknot and terminator pairing are shown by a dashed box (upper left).

(d-g) Fraction readthrough for 91C, 94C variants from (c) grouped as pseudoknot match (d), 29:91 mismatch (e), 26:94 mismatch (f), and 29:91/26:94 mismatch variants (g). Sequences are annotated as in (b). Red indicates a mismatch.

(h-k) Models for rescue of a 114:94 terminator mismatch (h) by a 29:91 (i) or 26:94 (j) pseudoknot mismatch to fold the terminator hairpin (k). Stars indicate mismatch mutations. All data in panels (b-g) are from n=2 independent replicates of each ZMP condition for the mutagenesis library depicted in a; 'Rep 1' and 'Rep 2' in panels (b) and (d-g) are annotated and heatmaps in panel (c) are an average. Replicates are compared in Supplementary Figure 7d.

Source data are available in the Northwestern University Arch Institutional Repository (<https://doi.org/10.21985/N2220T>).

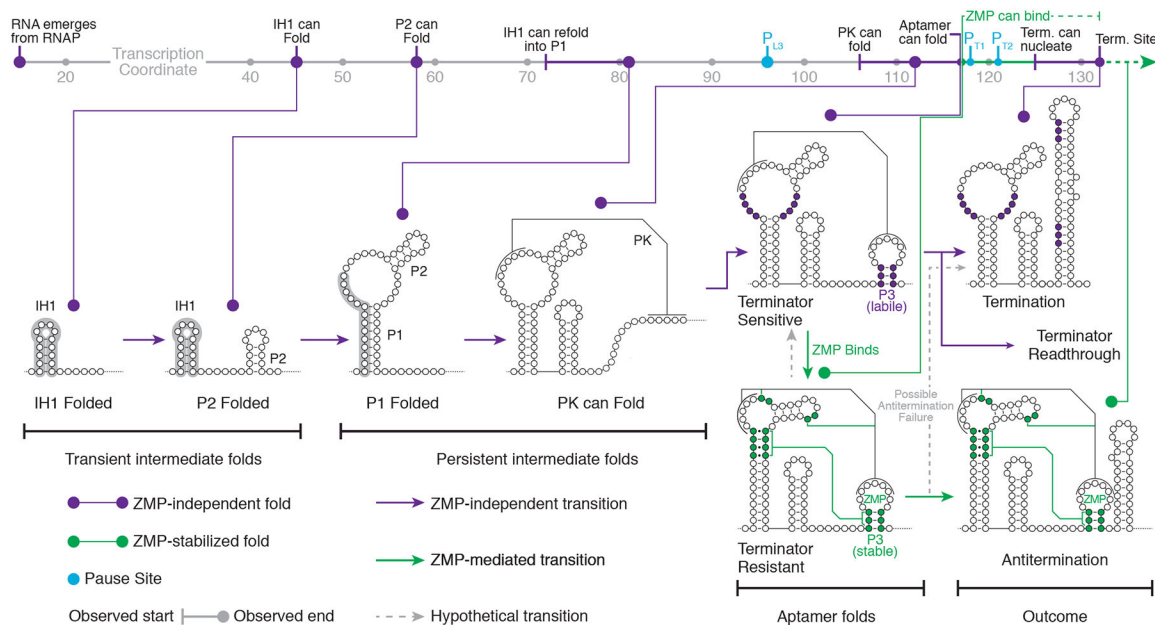


Figure 6. A model for *pfl* riboswitch folding

pfl 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; Grey indicates hypothetical transitions. Initial *pfl* aptamer folding comprises the formation of a transient intermediate hairpin (IH1) that rearranges to form P1. Upon folding of both the pseudoknot (PK) and P3, the *pfl* aptamer is competent to bind ZMP. While PK folding appears to occur before P3 folding in cotranscriptional SHAPE-Seq datasets (Figure 2), it is possible these events occur simultaneously or in the reverse order in unhindered transcription conditions. 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 *pfl* aptamer can become trapped in a stable fold that renders the P3 stem resistant to terminator nucleation, thereby driving antitermination.