nature chemical biology

ARTICLES

https://doi.org/10.1038/s41589-019-0382-7

A ligand-gated strand displacement mechanism for ZTP riboswitch transcription control

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

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.

The coupling of transcription and folding is ubiquitous within RNA biogenesis[1]. Nascent RNA folding is directed by the 5′ to 3′ polarity of transcription and the typically slower rate of nucleotide addition relative to base pair formation[2–4]. Consequently, cotranscriptional RNA folding favors local structures that can pose energetic barriers to the formation of long-range interactions[5–7]. The tendency of RNA molecules to enter such kinetic traps is thought to be the basis for gene regulation by riboswitches, noncoding 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 targets[11], diagnostic biosensors[12,13] and imaging tools[14].

Riboswitch architecture typically comprises a ligand-sensing aptamer and an 'expression platform' that together direct a regulatory outcome based on ligand occupancy[8]. These domains can overlap such that mutually exclusive aptamer and expression platform structures block or allow gene expression[8]. Cotranscriptional folding is crucial for riboswitches that regulate transcription because ligand recognition must occur within a limited window before expression platform folding[15]. While atomic-resolution structures of diverse riboswitch aptamers have provided a detailed understanding of ligand–aptamer complexes[9], the series of folding intermediates that mediate riboswitch function has only been described in a handful of cases[16–19]. 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 antitermination[20]. 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 that interact through a pseudoknot 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 J1/2 (ref. [20]) (Fig. 1a). Crystallographic studies of ligand-bound ZTP aptamers revealed extensive contacts between the aptamer subdomains, which could stabilize the aptamer against transcription termination[22–24] (Fig. 1b). However, the ZMP dependence of these interactions during transcription, and the precise antitermination mechanism, remains unclear.

Our analysis of the *Cbe 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*) RNA polymerase (RNAP) under limiting nucleoside 5′-triphosphate (NTP) conditions[20]. To evaluate *pfl* riboswitch control of *E. coli* 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 control[15], slower transcription at 100 µM NTPs reduced basal terminator readthrough and increased antitermination relative to faster transcription at 500 µM NTPs (Fig. 1c and Supplementary Fig. 1a). Conversely, inclusion of

[1]Department of Chemical and Biological Engineering, Northwestern University, Evanston, IL, USA. [2]Interdisciplinary Biological Sciences Graduate Program, Northwestern University, Evanston, IL, USA. [3]Center for Synthetic Biology, Northwestern University, Evanston, IL, USA. [4]Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, USA. *e-mail: eric.strobel@northwestern.edu; jblucks@northwestern.edu

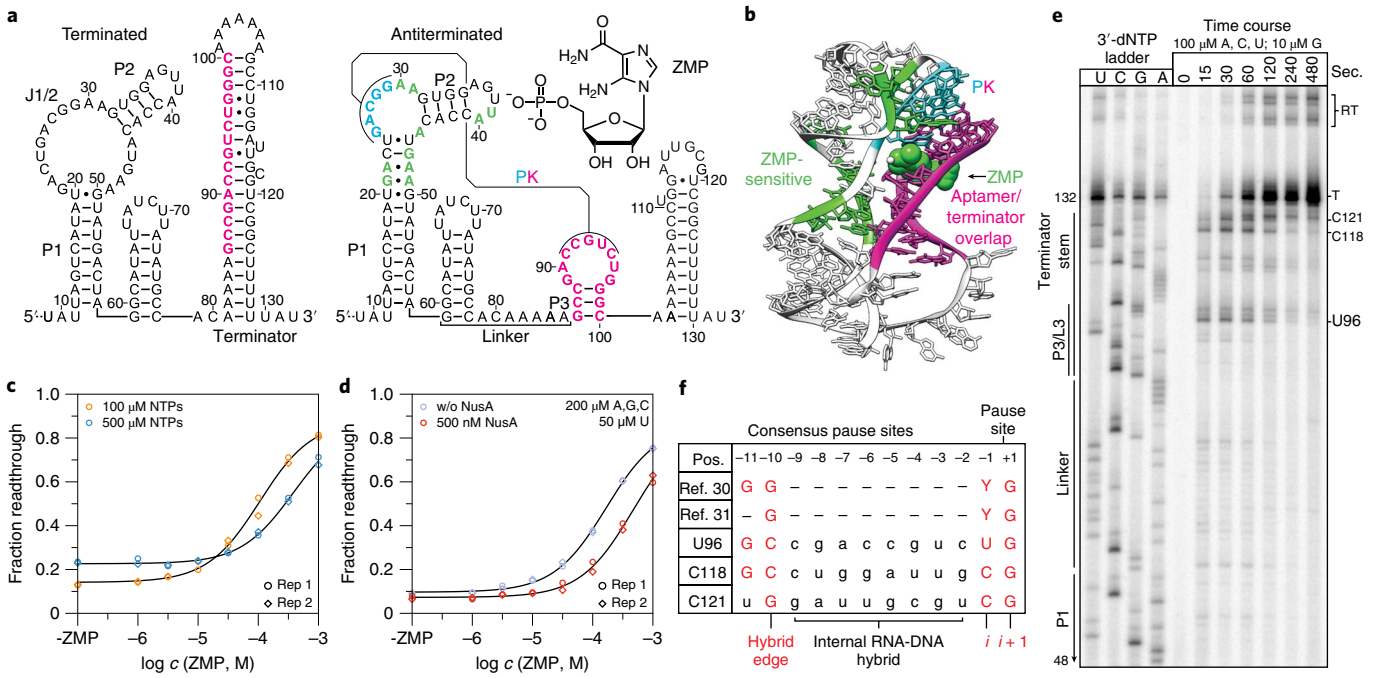

Fig. 1 | Overview and in vitro characterization of the *C. beijerinckii pfl* riboswitch. a, Secondary structure of the *Cbe pfl* riboswitch terminated and antiterminated folds[20,22–24]. Aptamer/terminator overlap is colored magenta, J1/2 pseudoknot nucleotides are cyan and ZMP-responsive nucleotides (see Fig. 2) are green. **b**, Crystal structure of the *Thermosinus carboxydivorans pfl* riboswitch (PDB 4ZNP)[24] highlighted to match the 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 µM and 500 µM NTPs (**c**) and with or without (w/o) 500 nM NusA (**d**). Representative gels are shown in Supplementary Fig. 1a,b. **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 µM and 1 mM ZMP data were obtained separately from values for ≤100 µM ZMP. In panel **e**, $n = 1$ and agrees with comparable measurements in the same and additional synchronization, ZMP and NusA conditions in Supplementary Fig. 1c,d. RT, readthrough; T, terminated; PK, pseudoknot; Sec, seconds; Pos, position. Uncropped source gels are shown in Supplementary Fig. 12. Source data are available online and in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N2220T).

E. coli NusA promoted termination[25] (Fig. 1d and 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 *E. coli* RNAP with 500 µM 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 *E. coli* RNAP may differ from *Cbe* RNAP in properties such as speed, pausing and nascent RNA interactions[5,27], our measurements describe a functional ZTP riboswitch.

Consensus pauses punctuate *pfl* riboswitch transcription. Transcription pausing[28] can facilitate riboswitch folding and ligand-binding kinetics[15,17,19] and can be sensitive to aptamer state[29]. Although pause-sequence recognition can be species specific[27], diverse multisubunit 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 to 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 *E. coli* RNAP across every template position[34]. 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 nucleotides 12–25 at transcript length ~45 when reduced reactivity in nucleotides 14–16 produces a 'low–high–low' reactivity pattern that is characteristic of a simple hairpin[18] (Fig. 2a,b). IH1 persists through to transcript length ~70, during which P2 folding is observed as increased reactivity at nucleotides 38–40 flanked by low reactivity at nucleotides 32–35 and 41–44 (Fig. 2a–c). From transcript lengths ~70–81, IH1 loop reactivity decreases as J1/2 (nucleotides 21–31) reactivity increases, suggesting rearrangement of IH1 to form P1 (Fig. 2a,c,e). Given the ~14 nucleotide footprint of *E. coli* RNAP on nascent RNA[35], 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

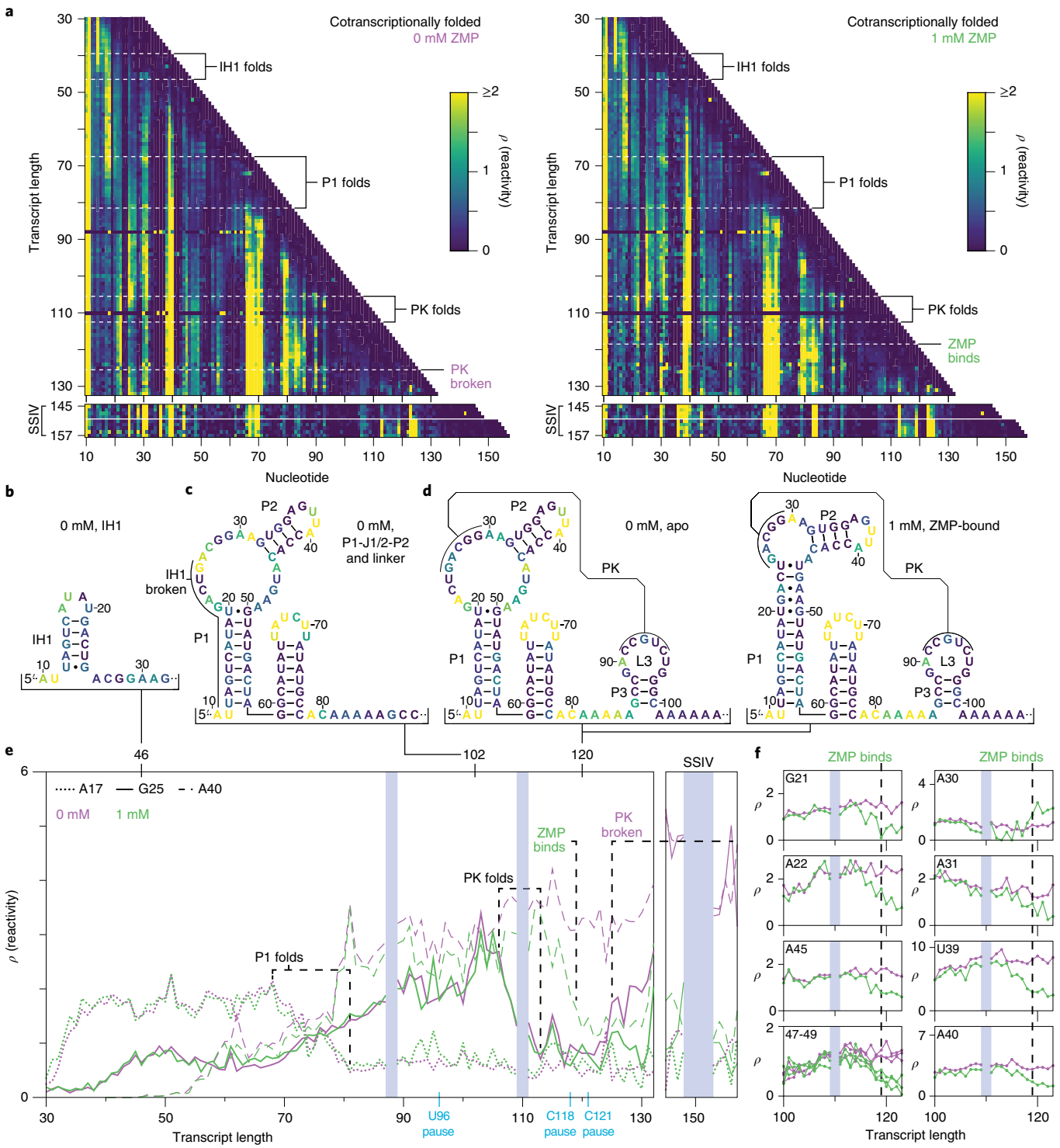

Fig. 2 | *C. beijerinckii pfl* riboswitch folding intermediates. a, Cotranscriptional SHAPE-seq reactivity matrices for the *Cbe pfl* riboswitch with 0 mM and 1 mM ZMP. The unstructured leader (nucleotides 1–9) is not shown. Data for transcripts 88 and 110 are absent due to ambiguous 3′ end mapping of sequencing reads. Reactivity spectra for transcripts 145–148 and 153–157 are from a separate targeted experiment using SSIV reverse transcriptase. Reactivity (ρ) is capped at a value of 2 for presentation of the heat map. **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 minimum free energy prediction informed by covariation[20] 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 (noncanonical 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 online and in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N2220T).

bacterial RNAPs[1], the persistence of IH1 for at least 35 nucleotide addition cycles suggests that it can form during uninterrupted transcription. Minimum free energy structure prediction[36] indicated that ~50% of 532 ZTP riboswitches from bacterial genomes[20] have the capacity for an intermediate structure that is as favorable as *Cbe* IH1 (Supplementary Fig. 3). Randomized control predictions suggest that 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 nucleotides 59–78, indicated by high reactivity from nucleotides 66–71 flanked by regions of low reactivity, and an adjacent 7 nucleotide unstructured region that comprise the linker between P1 and P3 (Fig. 2a,c). ZMP-independent pseudoknot folding is then observed as decreased reactivity at nucleotides 25–29 across transcript lengths 106–112, as L3 emerges from *E. coli* RNAP (Fig. 2a,e). In agreement, equilibrated RNA intermediates form the pseudoknot at transcript 95 when complete pairing between nucleotides 25–29 and 91–95 is possible (Supplementary Fig. 2). These data indicate that the pseudoknot can fold before P3, although the order of pseudoknot 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 *E. coli* RNAP (Fig. 2d–f). In agreement with this, 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 also disrupt and restore pseudoknot folding and ZMP binding, respectively (Supplementary Figs. 4 and 5). The requisite folding of P3 before the 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 measurements[20] and can be categorized as P1 or pseudoknot associated by comparing them with 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 nucleotides G21, A22, A45 and nucleotides 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 (refs. [22–24]) depends on ZMP binding. Further ZMP-dependent reactivity changes occur simultaneously at nucleotides A30, A31, U39 and A40, which directly contact or are proximal to pseudoknot base pairs[20,22–24] (Fig. 2f and 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 on stacking between A31 and the G29:C91 pseudoknot base pair[24]. Pseudoknot disruption renders the above nucleotides nonresponsive to ZMP and restoration of pseudoknot base pairs recovers detectable, but weaker, signatures of binding relative to the wild-type (WT) sequence (Supplementary Figs. 4–6). 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 nucleotides 25–29 in J1/2 from transcript lengths ~125–132, suggesting that pseudoknot disruption can begin as the 3′ terminator stem emerges from *E. coli* 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 nucleotide 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-nucleotide 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 biases[37], SSIV was less sensitive to adduct detection by complementary DNA 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 nucleotides G25, A31 and 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 mutations[38] and other in vitro transcription approaches that systematically perturb RNA transcripts[39], and is similar to a recent in-cell fluorescence-based genetic screen for riboswitch function[40]. 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 and 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

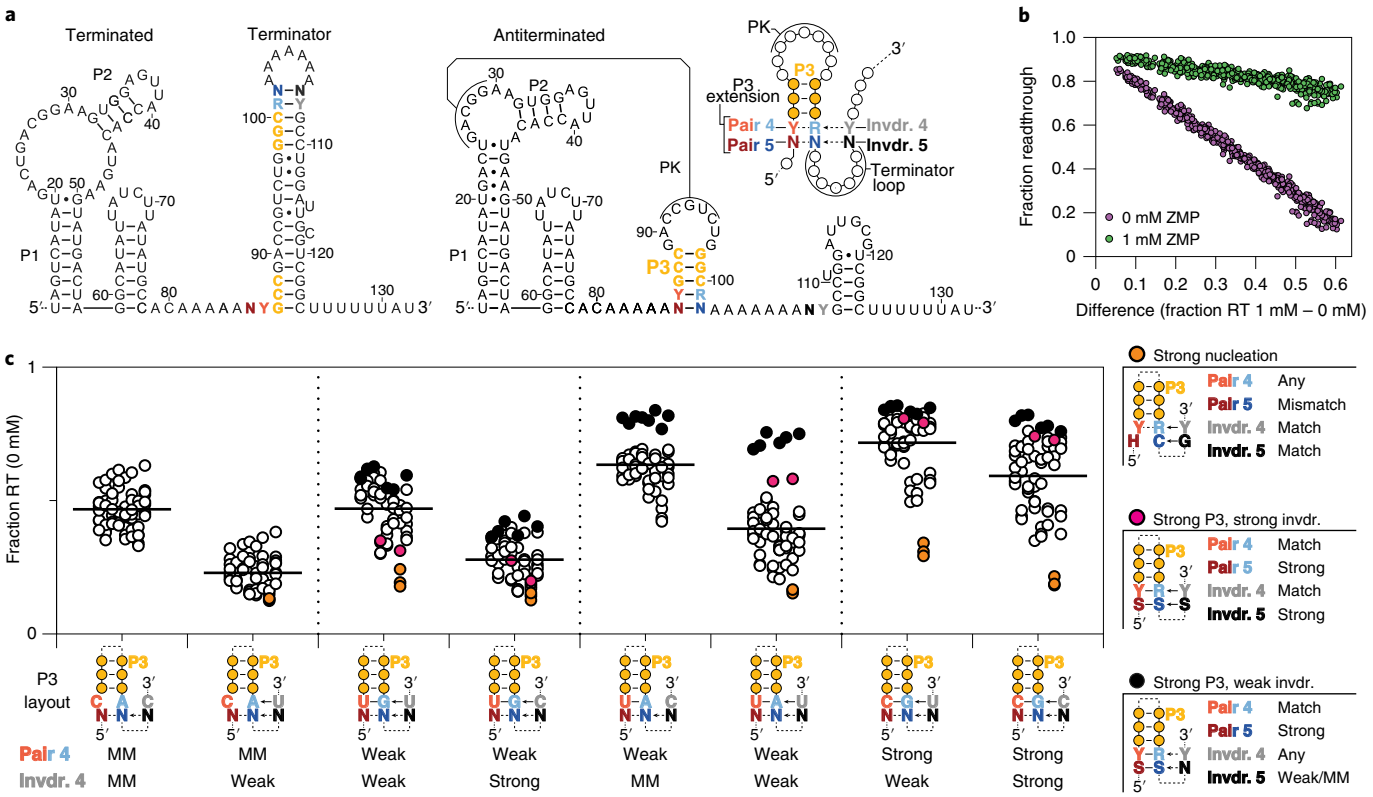

Fig. 3 | Mutagenesis of the *C. beijerinckii pfl* aptamer P3 stem. a, *Cbe pfl* riboswitch terminated and antiterminated secondary structures depicting P3 randomization scheme; randomized nucleotides are color coded. Insertion mutations are not counted in nucleotide numbering for consistency. The proposed 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 (1 mM – 0 mM). **c**, Fraction readthrough with 0 mM ZMP grouped by Pair 4 and Invdr 4 identity. Horizontal lines are group averages. P3 layout is colored as in **b**. Notable Pair 4/Invdr 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 Fig. 7c. Source data are available online and in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N2220T).

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 sequence[20] (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 readthroughs 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 and 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 the *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 was preserved[20] (Fig. 4a). These variants displayed a range of fraction readthroughs (1 mM – 0 mM 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

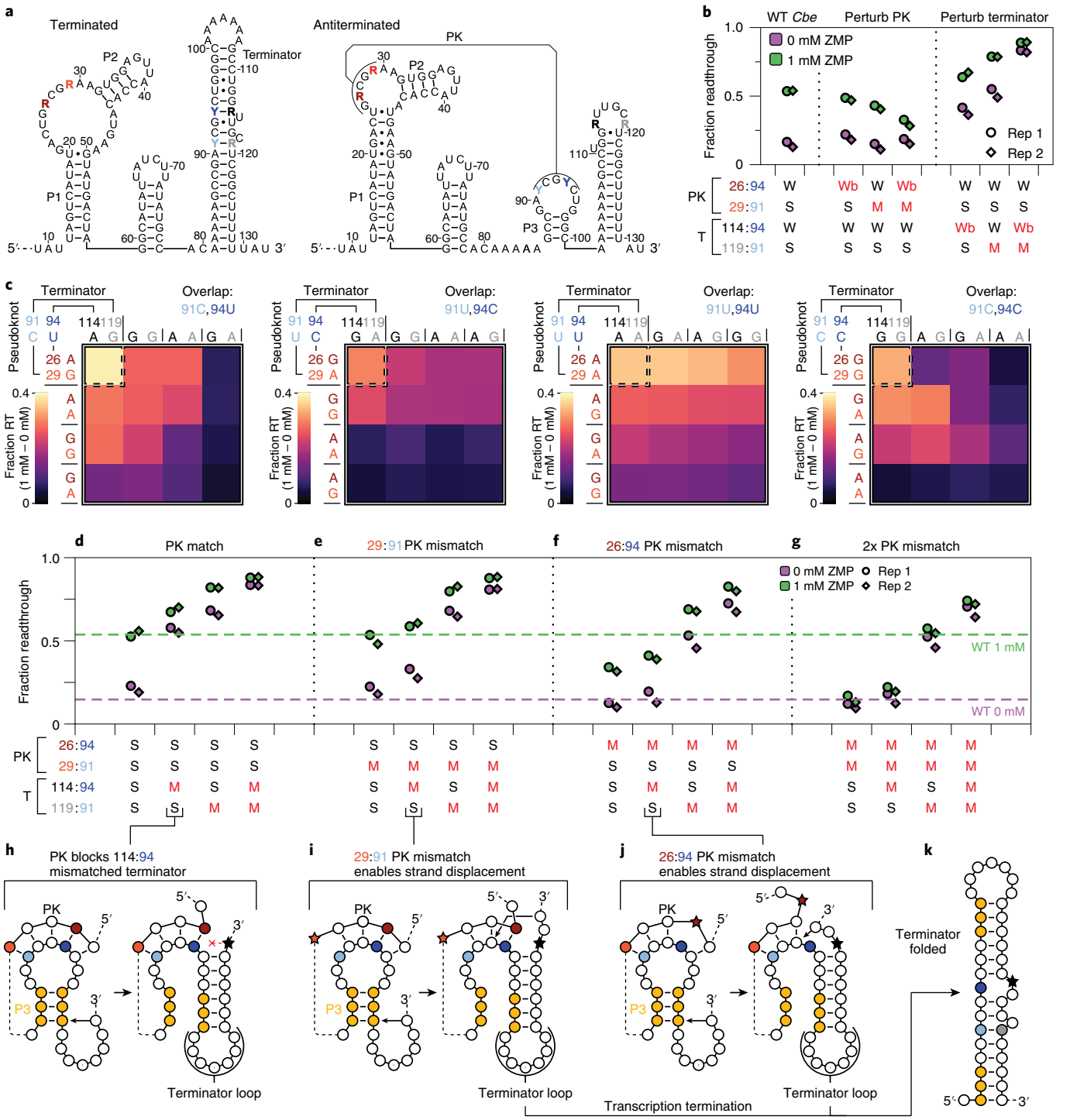

Fig. 4 | Mutagenesis of *C. beijerinckii pfl* aptamer pseudoknot and terminator base pairs. a, *Cbe pfl* riboswitch terminated and antiterminated secondary structures depicting pseudoknot randomization scheme; randomized nucleotides are color coded. R = A or G; Y = C or U. **b**, Fraction readthrough for select pseudoknot mutants with 0 mM or 1 mM ZMP. Variant pairing patterns in the pseudoknot and terminator 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 (nucleotides 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**. Replicates (Reps) 1 and 2 in panels **b**, **d**–**g** are annotated and heat maps in **c** are an average. Replicates are compared in Supplementary Fig. 7d. Source data are available online and in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N2220T).

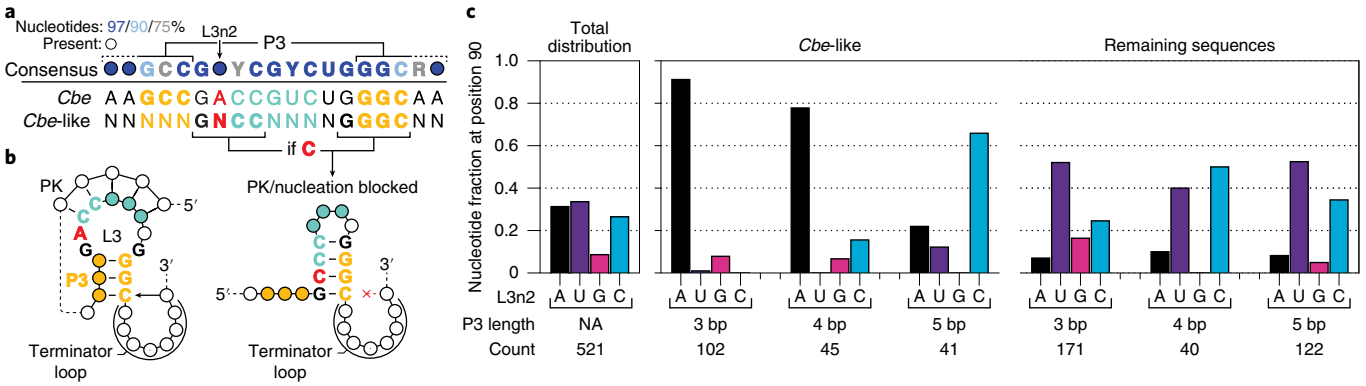

Fig. 5 | Classification of ZTP aptamer P3/L3 sequence and structure. a, Consensus, *Cbe* and *Cbe*-like sequences for the ZTP aptamer P3/L3 region. Defined and undefined consensus sequence positions are conserved by identity and presence, respectively. **b**, A model for how the A90C P3 misfold could interfere with pseudoknot and terminator folding in *Cbe*-like sequence contexts. Nucleotides are colored as in **a**. **c**, Nucleotide frequency at L3 nucleotide 2 for ZTP riboswitch sequences in aggregate and binned by presence of the '*Cbe*-like' sequence defined in **a** and the length of the P3 stem. Sequences with nonconserved insertions or a sequence alignment-determined P3 stem of <3 base pairs were not considered. NA, not applicable. Source data are available online and in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N2220T).

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 for 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, which 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 alternative pseudoknot (Supplementary Fig. 10c). Supporting this interpretation, N31 and N40 variants, which 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 by 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 for its identity[20], 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* nucleotide 90). Remarkably, *Cbe*-like P3/L3 sequences are enriched for A and depleted for C when P3 is 3 base pairs (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 folding[15], and these same requirements can be critical for translation control[19]. In this regard, the ZTP aptamer must overcome the challenge of folding its ligand-binding pocket from subdomains that are separated by a nonconserved linker[20]. 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 µM) 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, pausing[27] and RNA folding within the RNA exit channel[42], given the conservation of bacterial multisubunit RNAPs[43] 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

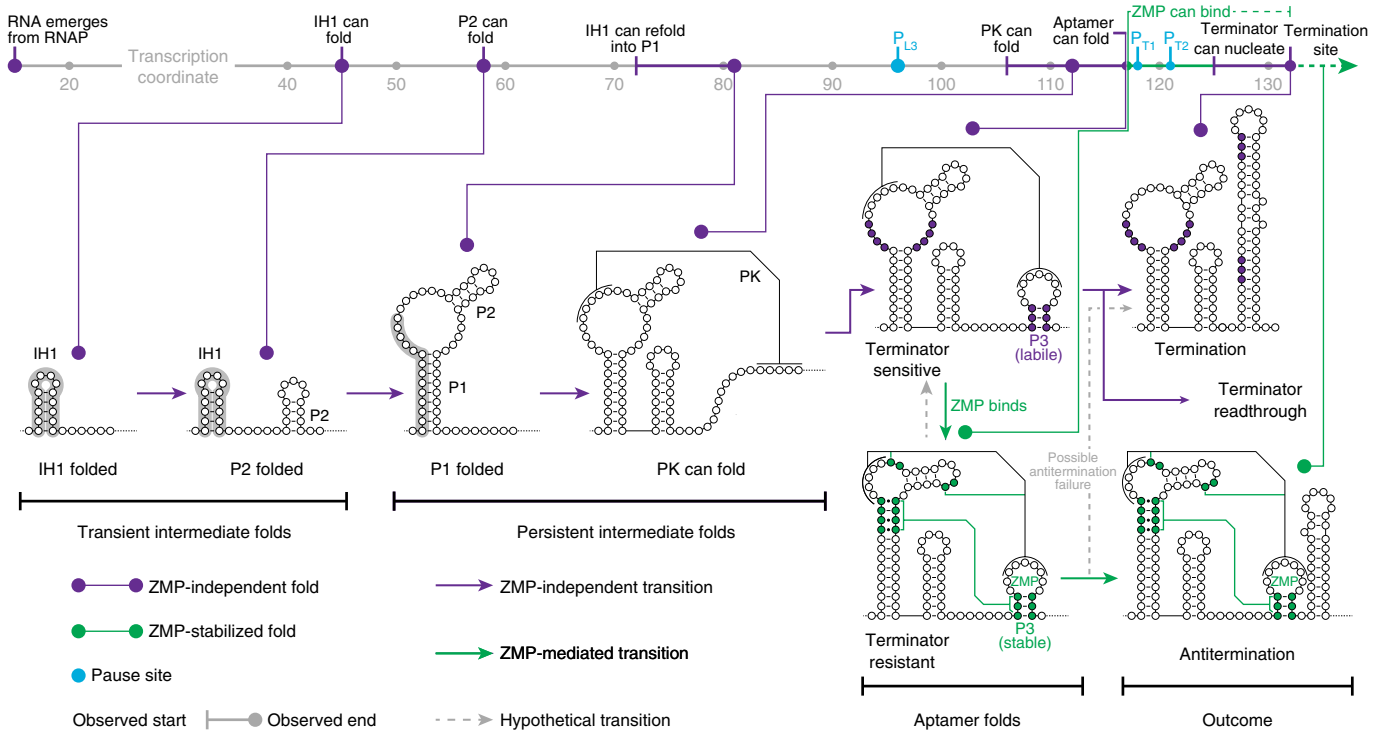

Fig. 6 | A model for *pfl* riboswitch folding. *pfl* riboswitch folding intermediates as observed by cotranscriptional SHAPE-seq are plotted by transcription coordinate. Purple indicates ZMP-independent folds and transitions; green indicates ZMP-stabilized folds or ZMP-mediated transitions; gray indicates hypothetical transitions. Initial *pfl* aptamer folding comprises the formation of a transient intermediate hairpin (IH1) that rearranges to form P1. On folding of both the pseudoknot and P3, the *pfl* aptamer is competent to bind ZMP. While pseudoknot folding appears to occur before P3 folding in cotranscriptional SHAPE-seq datasets (see Fig. 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 pseudoknot 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.

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 response[38].

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 (refs. [22–24]). Furthermore, ZMP recognition promotes noncanonical P1 base pairs that form a ribose zipper with P3 (refs. [22–24]) both in equilibrium[20] and cotranscriptionally (Fig. 2 and 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 riboswitch[46].

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 preQ$_1$ riboswitch[48] and RNA degradation control by the *GlmS* riboswitch[49].

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 structures[50]. The formation of transient folds[7] 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 it 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 noncoding RNAs. The ZTP aptamer also exhibits a context-dependent sequence preference, which 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 the pathways to fold that structure[3].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, statements of code and data availability and

associated accession codes are available at https://doi.org/10.1038/s41589-019-0382-7.

Received: 12 February 2019; Accepted: 22 August 2019;
Published online: 21 October 2019

References

1. Pan, T. & Sosnick, T. RNA folding during transcription. *Annu. Rev. Biophys. Biomol. Struct.* **35**, 161–175 (2006).
2. Zhang, J. & Landick, R. A two-way street: regulatory interplay between RNA polymerase and nascent RNA structure. *Trends Biochem. Sci.* **41**, 293–310 (2016).
3. Lai, D., Proctor, J. R. & Meyer, I. M. On the importance of cotranscriptional RNA structure formation. *RNA* **19**, 1461–1473 (2013).
4. Al-Hashimi, H. M. & Walter, N. G. RNA dynamics: it is about time. *Curr. Opin. Struct. Biol.* **18**, 321–329 (2008).
5. Pan, T., Artsimovitch, I., Fang, X. W., Landick, R. & Sosnick, T. R. Folding of a large ribozyme during transcription and the effect of the elongation factor NusA. *Proc. Natl Acad. Sci. USA* **96**, 9545–9550 (1999).
6. Heilman-Miller, S. L. & Woodson, S. A. Effect of transcription on folding of the *Tetrahymena* ribozyme. *RNA* **9**, 722–733 (2003).
7. Wong, T. N., Sosnick, T. R. & Pan, T. Folding of noncoding RNAs during transcription facilitated by pausing-induced nonnative structures. *Proc. Natl Acad. Sci. USA* **104**, 17995–18000 (2007).
8. Garst, A. D., Edwards, A. L. & Batey, R. T. Riboswitches: structures and mechanisms. *Cold Spring Harb. Perspect. Biol.* **3**, a003533 (2011).
9. McCown, P. J., Corbino, K. A., Stav, S., Sherlock, M. E. & Breaker, R. R. Riboswitch diversity and distribution. *RNA* **23**, 995–1011 (2017).
10. Nelson, J. W. & Breaker, R. R. The lost language of the RNA World. *Sci. Signal.* **10**, eaam8812 (2017).
11. Howe, J. A. et al. Selective small-molecule inhibition of an RNA structural element. *Nature* **526**, 672–677 (2015).
12. Kellenberger, C. A., Wilson, S. C., Sales-Lee, J. & Hammond, M. C. 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 (2013).
13. Porter, E. B., Polaski, J. T., Morck, M. M. & Batey, R. T. Recurrent RNA motifs as scaffolds for genetically encodable small-molecule biosensors. *Nat. Chem. Biol.* **13**, 295–301 (2017).
14. Braselmann, E. et al. A multicolor riboswitch-based platform for imaging of RNA in live mammalian cells. *Nat. Chem. Biol.* **14**, 964–971 (2018).
15. Wickiser, J. K., Winkler, W. C., Breaker, R. R. & Crothers, D. M. The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol. Cell* **18**, 49–60 (2005).
16. Frieda, K. L. & Block, S. M. Direct observation of cotranscriptional folding in an adenine riboswitch. *Science* **338**, 397–400 (2012).
17. Perdrizet, G. A. II, Artsimovitch, I., Furman, R., Sosnick, T. R. & Pan, T. Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc. Natl Acad. Sci. USA* **109**, 3323–3328 (2012).
18. Watters, K. E., Strobel, E. J., Yu, A. M., Lis, J. T. & Lucks, J. B. Cotranscriptional folding of a riboswitch at nucleotide resolution. *Nat. Struct. Mol. Biol.* **23**, 1124–1131 (2016).
19. Chauvier, A. et al. Transcriptional pausing at the translation start site operates as a critical checkpoint for riboswitch regulation. *Nat. Commun.* **8**, 13892 (2017).
20. Kim, P. B., Nelson, J. W. & Breaker, R. R. An ancient riboswitch class in bacteria regulates purine biosynthesis and one-carbon metabolism. *Mol. Cell* **57**, 317–328 (2015).
21. Bochner, B. R. & Ames, B. N. ZTP (5-amino 4-imidazole carboxamide riboside 5′-triphosphate): a proposed alarmone for 10-formyl-tetrahydrofolate deficiency. *Cell* **29**, 929–937 (1982).
22. Trausch, J. J., Marcano-Velazquez, J. G., Matyjasik, M. M. & Batey, R. T. Metal ion-mediated nucleobase recognition by the ZTP riboswitch. *Chem. Biol.* **22**, 829–837 (2015).
23. Jones, C. P. & Ferre-D'Amare, A. R. Recognition of the bacterial alarmone ZMP through long-distance association of two RNA subdomains. *Nat. Struct. Mol. Biol.* **22**, 679–685 (2015).
24. Ren, A., Rajashankar, K. R. & Patel, D. J. Global RNA fold and molecular recognition for a pfl riboswitch bound to ZMP, a master regulator of one-carbon metabolism. *Structure* **23**, 1375–1381 (2015).
25. Schmidt, M. C. & Chamberlin, M. J. nusA protein of *Escherichia coli* is an efficient transcription termination factor for certain terminator sites. *J. Mol. Biol.* **195**, 809–818 (1987).
26. Rohlman, C. E. & Matthews, R. G. Role of purine biosynthetic intermediates in response to folate stress in *Escherichia coli*. *J. Bacteriol.* **172**, 7200–7210 (1990).
27. Artsimovitch, I., Svetlov, V., Anthony, L., Burgess, R. R. & Landick, R. RNA polymerases from *Bacillus subtilis* and *Escherichia coli* differ in recognition of regulatory signals in vitro. *J. Bacteriol.* **182**, 6027–6035 (2000).
28. Artsimovitch, I. & Landick, R. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl Acad. Sci. USA* **97**, 7090–7095 (2000).
29. Widom, J. R. et al. Ligand modulates cross-coupling between riboswitch folding and transcriptional pausing. *Mol. Cell* **72**, 541–552.e6 (2018).
30. Larson, M. H. et al. A pause sequence enriched at translation start sites drives transcription dynamics in vivo. *Science* **344**, 1042–1047 (2014).
31. Vvedenskaya, I. O. et al. Interactions between RNA polymerase and the "core recognition element" counteract pausing. *Science* **344**, 1285–1289 (2014).
32. Strobel, E. J., Yu, A. M. & Lucks, J. B. High-throughput determination of RNA structures. *Nat. Rev. Genet.* **19**, 615–634 (2018).
33. Bevilacqua, P. C. & Assmann, S. M. Technique development for probing RNA structure in vivo and genome-wide. *Cold Spring Harb. Perspect. Biol.* **10**, a032250 (2018).
34. Strobel, E. J., Watters, K. E., Nedialkov, Y., Artsimovitch, I. & Lucks, J. B. Distributed biotin–streptavidin transcription roadblocks for mapping cotranscriptional RNA folding. *Nucleic Acids Res.* **45**, e109 (2017).
35. Komissarova, N. & Kashlev, M. Functional topography of nascent RNA in elongation intermediates of RNA polymerase. *Proc. Natl Acad. Sci. USA* **95**, 14699–14704 (1998).
36. Reuter, J. S. & Mathews, D. H. RNAstructure: software for RNA secondary structure prediction and analysis. *BMC Bioinformatics* **11**, 129 (2010).
37. Sexton, A. N., Wang, P. Y., Rutenberg-Schoenberg, M. & Simon, M. D. Interpreting reverse transcriptase termination and mutation events for greater insight into the chemical probing of RNA. *Biochemistry* **56**, 4713–4721 (2017).
38. Torgerson, C. D., Hiller, D. A., Stav, S. & Strobel, S. A. Gene regulation by a glycine riboswitch singlet uses a finely tuned energetic landscape for helical switching. *RNA* **24**, 1813–1827 (2018).
39. Vvedenskaya, I. O. et al. Massively systematic transcript end readout, "MASTER": transcription start site selection, transcriptional slippage, and transcript yields. *Mol. Cell* **60**, 953–965 (2015).
40. Polaski, J. T., Kletzien, O. A., Drogalis, L. K. & Batey, R. T. 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 (2018).
41. Larson, M. H., Greenleaf, W. J., Landick, R. & Block, S. M. Applied force reveals mechanistic and energetic details of transcription termination. *Cell* **132**, 971–982 (2008).
42. Hein, P. P. et al. RNA polymerase pausing and nascent-RNA structure formation are linked through clamp-domain movement. *Nat. Struct. Mol. Biol.* **21**, 794–802 (2014).
43. Geszvain, K. M. & Landick, R. in *The Bacterial Chromosome* (ed. Higgins, N. P.) Ch. 15 (American Society for Microbiology, 2005).
44. LeCuyer, K. A. & Crothers, D. M. Kinetics of an RNA conformational switch. *Proc. Natl Acad. Sci. USA* **91**, 3373–3377 (1994).
45. LeCuyer, K. A. & Crothers, D. M. The *Leptomonas collosoma* spliced leader RNA can switch between two alternate structural forms. *Biochemistry* **32**, 5301–5311 (1993).
46. Drogalis, L. K. & Batey, R. T. Requirements for efficient cotranscriptional regulatory switching in designed variants of the *Bacillus subtilis pbuE* adenine-responsive riboswitch. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/372573v1 (2018).
47. Zhao, B., Guffy, S. L., Williams, B. & Zhang, Q. An excited state underlies gene regulation of a transcriptional riboswitch. *Nat. Chem. Biol.* **13**, 968–974 (2017).
48. Liberman, J. A. et al. Structural analysis of a class III preQ$_1$ riboswitch reveals an aptamer distant from a ribosome-binding site regulated by fast dynamics. *Proc. Natl Acad. Sci. USA* **112**, E3485–E3494 (2015).
49. Savinov, A. & Block, S. M. Self-cleavage of the *glmS* ribozyme core is controlled by a fragile folding element. *Proc. Natl Acad. Sci. USA* **115**, 11976–11981 (2018).
50. Meyer, I. M. & Miklos, I. Co-transcriptional folding is encoded within RNA genes. *BMC Mol. Biol.* **5**, 10 (2004).

Acknowledgements

We thank R. Batey, C. Torgerson, S. Strobel, C. Jones and I. Artsimovitch for thoughtful discussions; J. Roberts (Cornell University) for providing *E. coli* NusA protein; J. Lis for use of facilities; J. Brink and S. Hockema for review of combinatorial mutagenesis alignment software; R. Breaker and K. Corbino for sharing a ZTP aptamer multiple sequence alignment; K. 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. 1DP2GM110838 to J.B.L.), Searle Funds at The Chicago Community Trust (to J.B.L.) and by the National Institute of General Medical Sciences (grant no. T32GM008382). 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

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

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41589-019-0382-7.

Correspondence and requests for materials should be addressed to E.J.S. or J.B.L.

Methods

DNA template preparation. Linear DNA templates were prepared by PCR amplification as described[51]. Briefly, five 100-µl reactions containing 82.25 µl of water, 10 µl Thermo Pol Buffer (New England Biolabs), 1.25 µl of 10 mM deoxyribonucleoside 5′-triphosphates (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; PerkinElmer) were added to a total of 100 nmol, such that ~1 biotin-11-dNTP is incorporated within the transcribed region of each DNA template[34]. DNA templates for combinatorial mutagenesis were prepared as above using oligonucleotides G and E (Supplementary Table 1) from gel-purified 'ultramer' oligonucleotides (Integrated DNA Technologies) after conversion to double-stranded DNA by eight PCR cycles containing 84.5 µl of water, 10 µl of 10× ThermoPol Buffer, 2 µl 10 mM dNTPs, 0.25 µl of 100 µM oligonucleotide F (forward primer; Supplementary Table 1), 0.25 µl of 100 µM oligonucleotide 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). For all DNA template preparations, 100-µl reactions were pooled and precipitated by adding 50 µl of 3 M 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 of water, 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).

Radiolabeled in vitro transcription. All radiolabeled single-round transcription reactions contained 10 nM DNA template and 0.016 U µl^{-1} *E. coli* RNA polymerase holoenzyme (New England Biolabs) in transcription buffer (20 mM Tris–HCl pH 8.0, 0.1 mM EDTA, 1 mM dithiothreitol (DTT) and 50 mM potassium chloride) and 0.2 mg ml^{-1} BSA (ref. [51]). When present, NusA (provided by J. Roberts, Cornell University) was included at 500 nM. ZMP (5-aminoimidazole-4-carboxamide-1-β-D-ribofuranosyl 5′-monophosphate; Sigma Aldrich) at a stock concentration of 50 mM in dimethylsulfoxide (DMSO) was added to variable concentration, with DMSO concentration fixed at 2% (v/v) in the final reaction. All reactions were 25 µl volume. 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) and 0.2 µCi µl^{-1} [α-^{32}P]UTP (PerkinElmer) at 37 °C for 10 min and initiated by adding $MgCl_2$ to 10 mM and rifampicin (Gold Biotechnology) to 10 µg ml^{-1} (ref. [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 position +15 of the DNA template relative to the transcription start site[20] by incubating reactions containing 2.5 µM ATP and GTP, 1.5 µM UTP, 0.2 µCi µl^{-1} [α-^{32}P]UTP and 10 mM $MgCl_2$ at 37 °C for 10 min before aliquoting to separate tubes containing 5× the desired concentration of NTPs, ZMP and rifampicin; reactions were incubated at 37 °C for 5 min before addition of 125 µl of stop solution. Synchronization for time-resolved single-round transcription[52] was performed either by stalling complexes at +15 and chasing with ATP, UTP and CTP to 100 µM, GTP to 10 µM (ref. [28]) and rifampicin to 10 µg ml^{-1} (Fig. 1e and Supplementary Fig. 1c) or by forming open promoter complexes with 100 µM of ATP and CTP, 50 µM of UTP and 10 µM of GTP and adding $MgCl_2$ to 10 mM and rifampicin to 10 µg ml^{-1} (Supplementary Fig. 1d). At each time point a single reaction volume was added to 125 µl of stop solution. When indicated, paused complexes were chased by adding GTP to 100 µM and incubating at 37 °C for 90 s. RNA sequencing ladders were generated by walking to +15 before adding 100 µM of NTPs and 100 µM of a chain terminating 3′-deoxyNTP (TriLink Biotechnologies); reactions proceeded for 5 min before addition of 125 µl of stop solution. All reactions were extracted by adding150 µl of phenol:chloroform:isoamyl alcohol (25:24:1), vortexing, centrifugation and collection of the aqueous phase, and then ethanol precipitated by adding 450 µl of 100% ethanol, 1.2 µL of GlycoBlue Coprecipitant (Thermo Fisher Scientific) and storing at −20 °C overnight. After centrifugation and removal of bulk and residual ethanol, precipitated RNA was resuspended in transcription loading dye (1× transcription buffer, 80% (v/v) formamide, 0.025% (w/v) bromophenol blue and xylene cyanol) and fractionated by denaturing urea–polyacrylamide gel electrophoresis (urea–PAGE) as described in 'Urea–PAGE and gel image quantification'.

In vitro transcription for cotranscriptional SHAPE-seq and mutagenesis experiments. Transcription reactions for nascent RNA structure probing and combinatorial mutagenesis were performed as previously described[34] by incubating 100 nM DNA template and 2 U of *E. coli* RNAP holoenzyme (New England Biolabs) in transcription buffer, 0.2 mg ml^{-1} BSA and 500 µM NTPs at 37 °C for 7.5 min to form open promoter complexes before adding streptavidin monomer (Promega) to 2.5 µM and continuing incubation for an additional 7.5 min; when streptavidin was not included, open complexes were formed for 10 min at 37 °C. When present, ZMP at a stock concentration of 50 mM in DMSO was added to a final concentration of 1 mM ZMP and 2% (v/v) DMSO; for samples without ZMP, DMSO was added to 2% (v/v). Transcription was initiated by adding $MgCl_2$ to 10 mM and rifampicin (Sigma Aldrich) to 10 µg ml^{-1} for a total reaction volume of 50 µl (cotranscriptional SHAPE-seq) or 25 µl (combinatorial mutagenesis and standard transcription). Transcription proceeded for 30 s. For cotranscriptional SHAPE-seq, chemical probing was then performed by splitting the sample into 25-µl aliquots and mixing with 2.78 µl of 400 mM benzoyl cyanide (BzCN; Pfaltz & Bauer) dissolved in anhydrous DMSO ((+) sample) or mixed with anhydrous DMSO ((−) sample) for ~2 s (ref. [53]). Transcription was stopped by adding 75 µl of TRIzol solution (Life Technologies) and RNAs were extracted according to the manufacturer's protocol. DNA template was degraded by incubation in 20 µl 1× DNase I buffer (New England Biolabs) containing 1 U DNase I (New England Biolabs) at 37 °C for 30 min. Water (30 µl) and 150 µl TRIzol were added and RNAs were extracted a second time. Depending on application, the resulting RNAs were then processed in one of several ways described in the sections 'Equilibrated RNA structure probing', 'Sequencing library preparation' and 'Urea–PAGE and gel image quantification'.

Equilibrated RNA structure probing. Transcription for equilibrium refolding experiments was performed as above, except that all reactions contained ZMP to promote stable distribution of elongation complexes across all 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 µl of water, equilibrium refolding was performed by denaturing at 95 °C for 2 min, snap cooling on ice for 1 min and adding transcription buffer, 500 µM of NTPs, 10 mM of $MgCl_2$ and either ZMP to 1 mM ZMP/2% DMSO or 2% DMSO before incubation at 37 °C for 20 min. SHAPE modification with BzCN was performed as described above before addition of 30 µl of water and 150 µl of TRIzol, extraction according to the manufacturer's protocol and resuspension in 10 µl of 10% DMSO.

Sequencing library preparation. Sequencing libraries for cotranscriptional SHAPE-seq were prepared either as previously described[54], or with a modified protocol that uses SSIV for reverse transcription. All combinatorial mutagenesis libraries were prepared using the modified SSIV protocol. For convenience, all protocol modifications are described 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 20× scale and purified by TRIzol extraction as described[54]. RNA 3′ ligation was performed by combining 10 µl extracted RNAs in 10% DMSO with 0.5 µl of SuperaseIN (Life Technologies), 6 µl of 50% PEG 8000, 2 µl of 10× T4 RNA Ligase Buffer (New England Biolabs), 1 µl of 2 µM 5′-adenylated RNA linker and mixing by pipetting. T4 RNA ligase 2, truncated KQ (0.5 µl) (New England Biolabs), was then added and the reaction was mixed again and incubated at 25 °C for 3 h.

Reverse transcription. Following linker ligation, RNAs were precipitated by adding 130 µl of RNase-free water, 15 µl of 3 M NaOAc pH 5.5, 1 µl of 20 mg ml^{-1} glycogen and 450 µl of 100% EtOH and stored at −80 °C for 30 min, centrifuged, washed once with 500 µl of 70% EtOH (v/v) and residual ethanol was removed. For SSIII reverse transcription, samples were resuspended in 10 µl of RNase-free water and 3 µl of 0.5 µM reverse transcription primer (oligonucleotide L, Supplementary Table 4), denatured at 95 °C for 2 min, incubated at 65 °C for 5 min, briefly centrifuged and placed on ice. SSIII reverse transcription master mix (7 µl; containing 4 µl of 5× First Strand Buffer (Life Technologies), 1 µl of 100 mM DTT, 1 µl of 10 mM dNTPs, 0.5 µl of RNase-free water and 0.5 µl of SSIII) was added and mixed before placing each sample at 45 °C and incubating at 45 °C for 1 min, 52 °C for 25 min and 65 °C for 5 min. For SSIV reverse transcription, RNAs were precipitated as above, resuspended in 9.5 µl of RNase-free water and 3 µl of reverse transcription primer (oligonucleotide L, Supplementary Table 4), denatured at 95 °C for 2 min, incubated at 65 °C for 5 min, briefly centrifuged and placed on ice. SSIV reverse transcription master mix (7.5 µl; containing 4 µl of 5× SSIV Buffer (Life Technologies), 1 µl of 100 mM DTT, 1 µl of 10 mM dNTPs, 0.5 µl of RNase OUT (Invitrogen) and 1 µl of SSIV) was added and mixed before placing each sample at 45 °C and incubating at 45 °C for 1 min, 52 °C for 25 min, 65 °C for 5 min and 80 °C for 10 min. Sodium hydroxide (1 µl of 4 M) was added and samples were heated at 95 °C for 5 min to hydrolyze RNA, partially neutralized by 2 µl of 1 M HCl and precipitated by adding 69 µl 100% EtOH and storing at −80 °C for 15 min, centrifugation for 15 min at 4 °C 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 described[54]. Briefly, dissolved cDNA was mixed with 3 µl of 10× 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 4) and 1 µl of CircLigase I (Epicentre), incubated at 60 °C for 2 h and 80 °C for 10 min. DNA was precipitated by adding 70 µl of nuclease-free water, 10 µl of 3 M NaOAc pH 5.5, 1 µl of 20 mg ml^{-1} glycogen and 300 µl of 100% EtOH and storing at −80 °C 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) according to the manufacturer's protocol and eluted with 20 µl of 1× TE buffer.

Quality analysis. Sequencing library quality analysis was performed as previously described[54] by generating fluorescently labeled double-stranded DNA 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 described[54]. Briefly, 3 µl of single-stranded DNA library (+) and (−) channels was separately mixed with 33.5 µl of nuclease-free water, 10 µl of 5× 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 65 °C 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 ten cycles. Following amplification, libraries were allowed to cool to 4 °C completely before the addition of 0.25 µl of ExoI (New England Biolabs) and incubation at 37 °C to degrade excess oligonucleotides. ExoI was heat inactivated by incubating at 80 °C 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 1× TE buffer and quantified using a Qubit 3.0 Fluorometer (Life Technologies). Molarity was estimated using the length distribution observed in the 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 v.1.0.1 as previously described[18], except that one mismatch was permitted during alignment following the observation that truncations at several nucleotides were enriched for a terminal mutation. For SSIV data we frequently observed a nontemplated T at full-length cDNA 3′ ends and we 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 v.3 (150-cycle). Libraries were loaded with a density of approximately 1,000 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. 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 (nucleotides 1–11) was not required because of elevated SSIV dropoff in this region and inclusion of these reads did not impact measurements (Supplementary Fig. 7f,g). Control analyses that omitted the poly(U) tract from alignment did not impact measurements (Supplementary Fig. 7f,g). Therefore, insertion mutants were aligned through two poly(U) tract nucleotides to permit usage of 150-cycle v.3 MiSeq Reagent Kits (Illumina). Single mismatches were permitted in Read1 beyond nucleotide 130 if unambiguous mapping as terminated or full length was possible. The 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). 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-^{32}P]UTP by dividing by U nucleotide count for each transcript. To normalize band intensity in experiments with staged transcription, the fraction [α-^{32}P]UTP/([α-^{32}P]UTP + UTP) was determined for the walk (P_{walk}) and chase with 100 µM or 500 µM of NTPs (P_{chase}) reaction phases and used to calculate labeling efficiency for each band with the equation ((P_{walk} × U nucleotides in walk) + (P_{chase} × U nucleotides in chase)); band intensity was divided by labeling efficiency. Labeling was virtually uniform because the probability of [α-^{32}P]UTP incorporation is high during the walk and low during the chase (~4.25% versus 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). In Fig. 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 a doublet; under the limiting GTP conditions the doublet bands most likely correspond to C121 and G122, which both precede a G.

Nonradiolabeled 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). Bands were normalized for RNA length by dividing band intensity by expected product length. Fraction readthrough was determined as above.

RNA structure prediction. RNA structure prediction was performed using the RNAstructure v.6.1 (ref. [36]) Fold command with default settings. For IH1 (Supplementary Fig. 3), WT sequences from previously identified ZTP riboswitches[20] were obtained from the RefSeq[55] entries used for their original identification and aligned using Infernal v.1.1.2 (ref. [56]). The aptamer segment used for structure prediction contains the sequence starting two nucleotides upstream of the longest P1 helices in the ZTP riboswitch family multiple sequence alignment[20] through to the last unstructured nucleotide 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 RNAstructure[36] Fold command to predict whether the 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. Wild-type 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 datasets match the WT length distribution.

L3n2 nucleotide frequency analysis. P3 hairpin sequences without insertions were binned by a match to the motif 'NNNNNGNCCNNNNGGGCNN' and then binned by the number of predicted contiguous P3 base pairs. Nucleotide frequency at L3 nucleotide 2 was then determined.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

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 5. SHAPE-seq Reactivity Spectra generated in this work have been deposited in the RNA Mapping Database[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 figures are available online and in the Northwestern University Arch Institutional Repository (https://doi.org/10.21985/N2220T). Uncropped gel images are shown in Supplementary Fig. 12. All other data that support the findings of this paper are available from the corresponding authors upon request.

Code availability

Spats v.1.0.1 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.

References

51. Strobel, E. J. & Roberts, J. W. Two transcription pause elements underlie a σ^{70}-dependent pause cycle. *Proc. Natl Acad. Sci. USA* **112**, E4374–E4380 (2015).
52. Landick, R., Wang, D. & Chan, C. L. Quantitative analysis of transcriptional pausing by *Escherichia coli* RNA polymerase: *his* leader pause site as paradigm. *Methods Enzymol.* **274**, 334–353 (1996).
53. Mortimer, S. A. & Weeks, K. M. A fast-acting reagent for accurate analysis of RNA secondary and tertiary structure by SHAPE chemistry. *J. Am. Chem. Soc.* **129**, 4144–4145 (2007).
54. Watters, K. E. et al. Characterizing RNA structures in vitro and in vivo with selective 2′-hydroxyl acylation analyzed by primer extension sequencing (SHAPE-Seq). *Methods* **103**, 34–48 (2016).
55. Pruitt, K. D., Tatusova, T. & Maglott, D. R. NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **35**, D61–D65 (2007).
56. Nawrocki, E. P. & Eddy, S. R. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**, 2933–2935 (2013).
57. Cordero, P., Lucks, J. B. & Das, R. An RNA mapping database for curating RNA structure mapping experiments. *Bioinformatics* **28**, 3006–3008 (2012).

nature research

Corresponding author(s): Eric Strobel
Julius Lucks

Last updated by author(s): Sep 30, 2019

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☒	☐ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐ A description of all covariates tested
☒	☐ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☒	☐ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☒	☐ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	Not Applicable
Data analysis	SHAPE-Seq data were analyzed by Spats v1.0.1. Multiple sequence alignment was performed using INFERNAL v1.1.2. RNA structure prediction was performed using RNAstructure v6.1. All custom software used for combinatorial mutagenesis data analysis, RNA structure prediction, and sequence analysis of riboswitch elements is freely available at https://github.com/LucksLab/LucksLab_Publications/tree/master/Strobel_ZTP_Riboswitch.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on 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 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. 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.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Individual contranscriptional SHAPE-Seq experiments inherently produce multiple measurements for each RNA folding intermediate because structures tend to persist across tens of transcripts. Similarly, combinatorial mutagenesis experiments achieve large sample sizes by exhaustively testing RNA sequence combinations.
Data exclusions	No samples were excluded from analyses
Replication	Cotranscriptional SHAPE-Seq for the WT ZTP riboswitch was replicated in n=3 independent replicates for each condition, and targeted cotranscriptional SHAPE-Seq for the WT ZTP riboswitch with SSIV RT was replicated in n= 2 independent replicates; the correlation of measurements across replicates is plotted in Supplementary Figure 11. Equilibrium and mutant cotranscriptional SHAPE-Seq measurements performed once. All combinatorial mutagenesis experiments are n=2 independent replicates for each condition and the correlation of replicate pairs is plotted in Supplementary Figure 7. Targeted measurements to confirm trends observed in combinatorial mutagenesis screens were performed once. Dose-response curves were n=2 independent replicates. Each time-resolved single-round in vitro transcription experiment was performed once, but trends were reproduced across four conditions and two transcription synchronization strategies.
Randomization	Randomization is not relevant to our study because our biochemical reactions are handled uniformly except for the specific condition being tested
Blinding	Blinding is not relevant to our study because the experimental measurements are automated and deterministic, and are therefore not subject to human biases.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging