

Regulation of the thiamine pyrophosphate (TPP)-sensing riboswitch in *NMT1* mRNA from *Neurospora crassa*

Sha Gong¹ , Chengyi Du¹ and Yanli Wang²

¹ College of Mathematics and Physics, Huanggang Normal University, China

² Department of Physics, Wuhan University, China

Correspondence

S. Gong, College of Mathematics and Physics, Huanggang Normal University, Huanggang, Hubei 438000, China
 Tel: 0713-8835186
 E-mail: shagong@whu.edu.cn

(Received 4 August 2019, revised 16 October 2019, accepted 22 October 2019, available online 13 November 2019)

doi:10.1002/1873-3468.13654

Edited by Michael Ibba

The expression of *Neurospora crassa* *NMT1* involved in thiamine pyrophosphate (TPP) metabolism is regulated at the level of mRNA splicing by a TPP-sensing riboswitch within the precursor *NMT1* mRNA. Here, using the systematic helix-based computational method, we investigated the regulation of this riboswitch. We find that the function of the riboswitch does not depend on the transcription process. Whether TPP is present or not, the riboswitch predominately folds into the ON state, while the OFF state aptamer structure does not appear during transcription. Since the transition from the ON state to the aptamer structure is extremely slow, TPP may interact with the RNA before full formation of the aptamer structure, promoting the switch flipping. The potential to fully form helix P0 of the ON state is necessary to restore ligand-dependent gene control by the riboswitch.

Keywords: cotranscriptional folding; gene regulation; TPP riboswitch

Riboswitches, as self-regulatory elements, are involved in regulating downstream gene expression through conformational changes driven by sensing specific metabolites or other physical signals [1–6]. Generally, a riboswitch is composed of two domains: a conserved aptamer domain responsible for ligand binding and an expression platform [7–9]. Upon ligand binding, the aptamer is stabilized, thereby altering base pair arrangements to expose or sequester the regulatory element in the expression platform. The structural arrangement of the expression platform directly modulates the expression of downstream coding genes at levels of transcription [10–12], translation [3,13], and RNA splicing [14,15]. As genetic control elements, riboswitches can be characterized by a signal-dependent RNA structural shift between two distinct functional states, that is, the ligand-bound state and ligand-unbound state. One of the two states serves as the genetic OFF state, which may contain a transcription terminator or a paired region covering the

translation initiation site. The switch flips when forming the alternative structure that destroys these regulatory elements. For the riboswitches acting at the level of RNA splicing, the main difference of the two functional structures is the structural flexibility of the relevant splice site. Depending on the concentration of cellular metabolites, one of the two functional structures is adopted to regulate downstream gene expression.

Thiamine pyrophosphate (TPP)-sensing riboswitches are the most widespread riboswitch class known in bacteria. They have also been identified in fungi [16], archaea, and plants [16–18], and are among the earliest discovered riboswitch representatives. They can regulate gene expression involved in the biosynthesis and transport of the coenzyme thiamine and its phosphorylated derivatives [19–21], by controlling messenger RNA translation [18], and splicing or processing [19,22,23]. The TPP riboswitch in *NMT1* mRNA from *Neurospora crassa* (Fig. 1) regulates the expression of

Abbreviations

N. crassa, *Neurospora crassa*; TPP, thiamine pyrophosphate.

a protein involved in TPP metabolism at the splicing level [23]. The aptamer domain forms a tuning fork architecture to bind TPP in which the prongs are formed by two stacking helices and arranged *via* a central three-way junction with stem P1 [22,24,25]. The docking of the ligand to the aptamer proceeds in two steps, from a weakly bound state to a strongly folded state [26].

The TPP riboswitch modulates gene expression through structural changes in the 5' splice site triggered by ligand-induced conformational rearrangements [23]. This structural difference will directly lead to different spliced products, which can repress or induce *NMT1* gene expression. When TPP concentration is high, TPP binding to the aptamer structure (OFF state) increases the structural flexibility near the 5' splice site. The effect of the structural change yields

the long-spliced mRNAs with the short upstream open reading frame, which competes with the translation of the main open reading frame and represses *NMT1* gene expression. In the absence of TPP, the riboswitch adopts an alternative structure (ON state) that occludes the 5' splice site, leading to a short-spliced mRNA which induces the expression of *NMT1* gene. Besides TPP, the riboswitch with mutated P3 helix has been found to prevent gene expression even when the ligand is absent [23].

The high-resolution structure of the bound aptamer has been solved, while structural information of ON state is only inferred from the experiments that it occludes the 5' splice site [23]. *In vivo*, nascent RNAs fold as soon as they are transcribed [27]. This sequential process, with associated structures formed within seconds, poses a great challenge to detect these structures, especially for the transcription of the long RNA chain. As the information of conformation formation and transition during the transcription are unclear, the structure of ON state and regulation mechanism of the TPP riboswitch have not been well characterized. Most previous literature focuses on the products of the transcript alone or the ligand binding kinetics [23,28–31], while little has been done to fully characterize how the riboswitch acts in cells, which is the key question for understanding and manipulating riboswitch function. By using the recently developed systematic helix-based computational method [32,33], we studied the regulatory behavior of the TPP riboswitch from *N. crassa* and a mutated construct that was found to prevent gene expression without the ligand. The results suggested that during transcription without the ligand, the riboswitch predominantly folds into the lower-energy ON state while the aptamer structure is not formed. The main cotranscriptional folding pathway is the same as that when the ligand is present. Similarly to other thermodynamically controlled riboswitches [34,35], function of the riboswitch is not sensitive to the transcription context as the ligand binds post-transcriptionally. However, since the transition from the ON state to the aptamer structure (OFF state) involves large conformational rearrangements, the transition rate is much slower than the splice rate [36]. In order to function as a ligand-responsive genetic switch, the ligand may interact with the riboswitch before the full formation of the aptamer and aid the conformation switching.

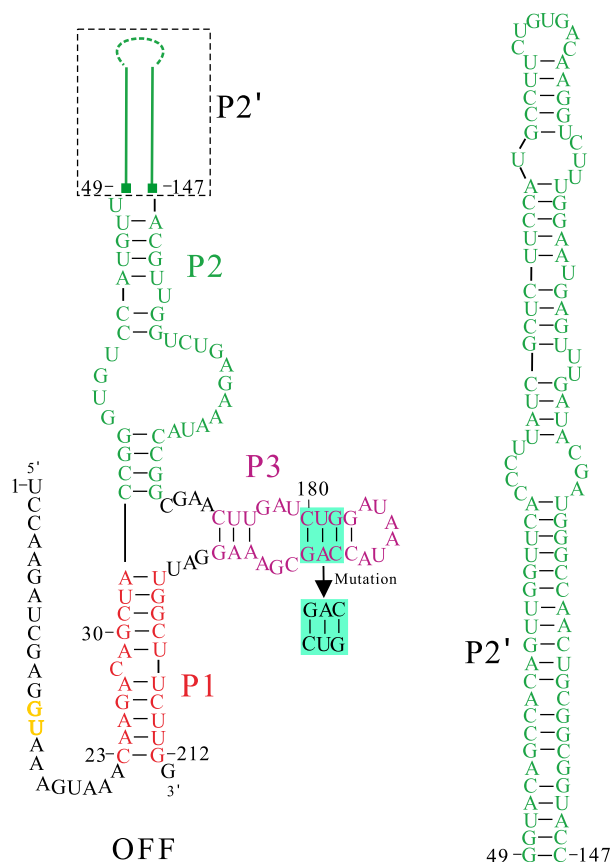


Fig. 1. The structure model of OFF state of the TPP riboswitch in the *NMT1* gene from *N. crassa*. Nucleotides within helices P1, P2, and P3 and the 5' splice site are colored differently: P1 (red), P2 (green), P3 (purple), and the 5' splice site (yellow). The green line denotes the upper part of long stem P2, which is characterized detailedly on the right (denoted by helix P2'). The green squares denote the bottom two paired nucleotides in helix P2'. The nucleotides in the box are the mutated P3 region.

Materials and methods

The approach to calculate cotranscriptional folding kinetics is based on the theory to predict free-folding kinetics. Thus,

in this section, we first introduce these two theories to calculate folding kinetics, followed by the transition node approximation, which is employed to explore the cotranscriptional folding of the TPP riboswitch.

Free-folding kinetics

The conformation space of a given RNA sequence consists of all possible secondary structures by enumeration. The nearest-neighbor model is used to calculate free energy of each structure [37,38]. The population kinetics at time t can be described by the master equation: $d\mathbf{p}(t)/dt = \mathbf{M} \bullet \mathbf{p}(t)$ [39], where $\mathbf{p}(t)$ denotes the vector of the population distribution $\{p_1(t), p_2(t), \dots, p_i(t)\}$ and \mathbf{M} is the rate matrix with elements $M_{ij} = k_{i \rightarrow j} (i \neq j)$ and $M_{ii} = -\sum k_{i \rightarrow j}$. $p_i(t)$ is the population of state i at time t ; $k_{i \rightarrow j}$ and $k_{j \rightarrow i}^{\neq}$ are the rate constants for the respective transitions. For $t > 0$, the master equation yields the population kinetics [39,40]:

$$\mathbf{p}(t) = \sum_{m=1} C_m \mathbf{n}_m e^{-\lambda_m t} \quad (1)$$

where \mathbf{n}_m and $-\lambda_m$ are the m -th eigenvector and eigenvalue of the rate matrix, respectively, and the coefficient C_m is determined by the initial conditions.

Basically, transition between different states is achieved by the formation or disruption of a stack and the respective rate constants can be easily calculated [39,41]. However, instead of stack, helices are used as building blocks, which reduces the conformation space for the calculation of the overall folding kinetics [32,42,43]. In the helix-based model [32,39], a basic move is addition or deletion of a helix or an exchange between two helices.

The helix formation is the most probable through the zipping pathway. The zipping rate k_f along the $1 \rightarrow 2 \rightarrow 3$ pathway shown in Fig. 2A can be calculated as [39]:

$$\begin{aligned} k_f &= k_{1 \rightarrow 2} K_1 \left(1 - K_2' K_1' \sum_{n=0}^{\infty} (K_2' K_1')^n \right) \\ &= k_{1 \rightarrow 2} K_1 \left(1 - K_2' K_1' \frac{1}{1 - K_2' K_1'} \right) \end{aligned} \quad (2)$$

where $K_1(K_2)$ and $K_1'(K_2')$ are the forward and reverse probability for state 2 (3),

$$\begin{aligned} K_1 &= \frac{k_{2 \rightarrow 3}}{k_{2 \rightarrow 3} + k_{2 \rightarrow 1} + k_{2 \rightarrow 4}}, \quad K_1' = \frac{k_{2 \rightarrow 1}}{k_{2 \rightarrow 3} + k_{2 \rightarrow 1} + k_{2 \rightarrow 4}}, \\ K_2 &= \frac{k_{3 \rightarrow 5} + k_{3 \rightarrow 6}}{k_{3 \rightarrow 2} + k_{3 \rightarrow 5} + k_{3 \rightarrow 6}}, \quad K_2' = \frac{k_{3 \rightarrow 2}}{k_{3 \rightarrow 2} + k_{3 \rightarrow 5} + k_{3 \rightarrow 6}}. \end{aligned}$$

The rate k_F for formation of a helix is the sum of the rates along all zipping pathways with different nucleation base pairs. The rate for deletion of the helix is calculated from the detailed balance condition: $k_U = k_F e^{-\Delta G/k_B T}$, where ΔG is the folding free energy of the helix.

For two incompatible helices, a basic move is the helix exchange transition, which can occur through the

unfolding–refolding pathway and the tunneling pathway [39]. Compared to the unfolding–refolding pathway (Fig. 2B), the transition between the two helices is mainly through the tunneling pathway with a lower barrier [39]. For the tunneling pathway, after the nucleotides required for helix H_2 nucleation are released by partially disruption of helix H_1 , in each subsequent step, disruption of a stack in helix H_1 is followed by the formation of a stack in helix H_2 . The transition rates between helices H_1 and H_2 through tunneling pathway can be calculated as [39]:

$$\begin{aligned} k_{H_1 \rightarrow H_2} &= \frac{\prod_i^n k_i}{\sum_{j=0}^{n-1} (\prod_{i=1}^j k_i' \prod_{m=j+2}^n k_m)}, \quad k_{H_1 \rightarrow H_2} \\ &= k_{H_1 \rightarrow H_2} e^{-\Delta G_{H_1 H_2}/k_B T} \end{aligned} \quad (3)$$

where k_n and k_n' are the rate constants for the processes of formation (disruption) and disruption (formation) of a base stack in $H_1(H_2)$, respectively. $\Delta G_{H_1 H_2}$ is the free energy difference between the two helices.

Cotranscriptional folding kinetics

In the theory [32], the whole transcription process is divided into different transcriptional steps. If the transcription rate is v nucleotides per second, the time window for each transcriptional step will be seconds. Within the time window of each step, the relaxation of the population distribution of the chain can be described in the same manner as the free-folding kinetics by using Eq. 1. However, here the initial population at the each step is determined by the final population distribution of the previous step (Eqs. 4) according to the structure types listed in Fig. 2C [32]:

$$\begin{aligned} p(N)_{\text{begin}} &= p(N-1)_{\text{end}} \text{ for } a, b, \text{ and} \\ c; \quad p(N)_{\text{begin}} &= 0 \text{ for } d \end{aligned} \quad (4)$$

At the beginning of the first step, the initial population of the open chain is 1. For consecutive steps, the folding results of the previous step turn into the initial conditions of the next step. Applying this method from the first step to the end of transcription, the main transcription pathways can be obtained by computing the folding kinetics of the RNA chain during transcription.

Transition node approximation

For the long RNA chain, such as the TPP riboswitch, the transition node approximation is necessarily employed to further decrease the number of the calculated conformations [33]. The basic idea of this approximation is to discard some newly formed states in each step. Here, we still used the N -th step as an example to demonstrate this approximation. The conformation space at N -th is divided into four ensembles (Fig. 2C), according to their free energies and the initial population distribution [33,44]. For

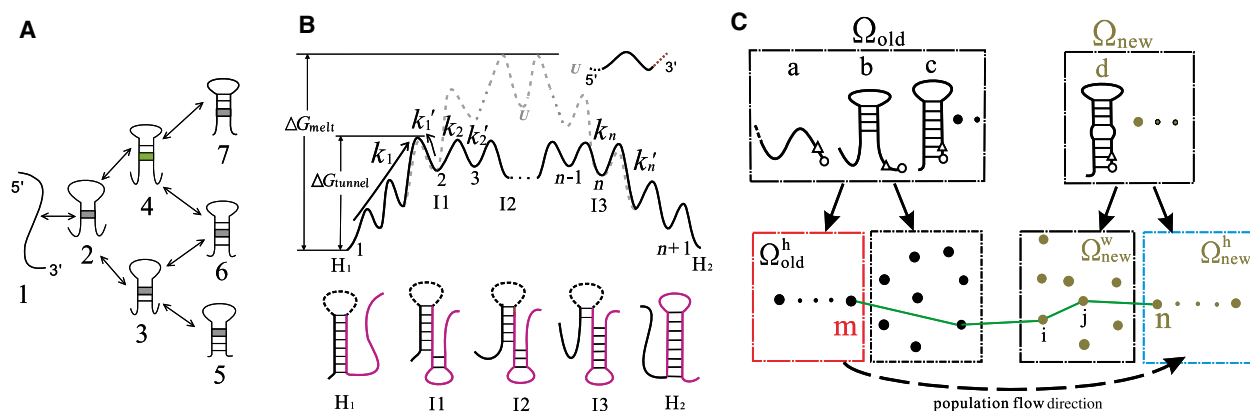


Fig. 2. (A) Formation of a helix along the zipping pathways. The first closing stack is denoted by the green color. (B) The free energy landscape for the transition between helices H_1 and H_2 . Solid line: the tunneling pathway. Dotted line: the unfolding-refolding pathway. (C) The population lineage and transition node approximation at the N -th step. The triangle and the circle denote the last nucleotide released by RNA polymerase at steps $N-1$ and N , respectively. The conformation space at the N -th step can be divided into two ensembles: Ω_{old} and Ω_{new} based on the initial population. Ω_{old}^h : a subensemble of Ω_{old} , in which the initial population of each state is larger than 0.03. Ω_{new}^h : a subensemble of Ω_{new} , in which each state has the free energy $\Delta G \leq (\min E + 3) \text{ kcal} \cdot \text{mol}^{-1}$, where $\min E$ is the free energy of the most unstable state in Ω_{old}^h . Ω_{new}^w : an ensemble of all rest states of Ω_{new} . The green line shows the assumed transition pathway between state m and state n , whereas the two intermediate states are denoted by i and j .

newly formed states, only those in Ω_{new}^h are likely to have the population accumulation due to their stabilities. Although states in Ω_{new}^w are impossible to have population aggregation, some may be on the main pathways for population flow from Ω_{old}^h to Ω_{new}^h . Therefore, the reduced N -nt conformation ensemble should include Ω_{old} , Ω_{new}^h , and parts of Ω_{new}^w .

Considering transition rates and free energies of intermediate states (nodes), only the pathways through which states in Ω_{old}^h can transit to states in Ω_{new}^h within eight nodes with corresponding transition rates greater than 0.0066 s^{-1} are taken into account in the approximation [33]. If no such transition pathway is remained, the criteria are relaxed to nine nodes or more to avoid isolate states in the transition network. After searching all possible transitions, the states in Ω_{new}^w that locate in saved transition pathways, together with Ω_{new}^h and Ω_{old} , comprise the conformation ensemble for subsequent calculations.

Results

Cotranscriptional folding behavior of the TPP riboswitch

The typical transcription speed in nature is around $50 \text{ nt} \cdot \text{s}^{-1}$ for bacteria [45]. Therefore, to investigate the regulatory behavior of the TPP riboswitch, we first calculated the population kinetics of the riboswitch during the transcription at an elongation rate of $50 \text{ nt} \cdot \text{s}^{-1}$ (Fig. 3). In the absence of TPP, the nascent RNA chain folds into a series of discrete intermediate states involving 11 main

stages [(1) to (11)] as the chain grows. (1): As the first 35 nucleotides are released, the open RNA chain (C0) begins to fold into the hairpin structure C1. (2): From step 43, structure C1 begins to be replaced by structure C2 with a long paired region. (3): A newly formed hairpin is directly added to structure C2 and structure C3 is formed at the 57th step. (4): As another new helix begins to be nucleated from step 66, almost all the riboswitches are populated at structure C4. (5): At the 92nd step, most of the riboswitches transit from C4 to the three-way junction structure C5 by closing the nonlocal helix. (6): With nucleation and elongation of the long helix P2', C5 begins to fold into structure C6 from the 112nd step. This structure, which is the most stable at step 123 ($\Delta G_{C6} = -31.64 \text{ kcal} \cdot \text{mol}^{-1}$), occupies most of the population from steps 115 to 132. (7): With elongation of helix P2', the middle helix in C6 is invaded by helix P2' and structure C6 begins to transit to C7 from step 132. By the 147th step, helix P2' is fully formed and C7 occupies about 86% of the population. (8): From step 155, C7 begins to fold into C8 by closing the bottom nonlocal helix. (9): With the formation of helix P5 at the 169th step, structure C8 folds into structure C9. (10): From about step 176, C9 is quickly replaced by structures C10 and C11. Structure C11 is fully formed at step 184 ($\Delta G_{C11} = -70.85 \text{ kcal} \cdot \text{mol}^{-1}$) with about 69% population, while structure C10 ($\Delta G_{C10} = -62.61 \text{ kcal} \cdot \text{mol}^{-1}$) occupies about 22% population. As two compatible hairpins (a, c) are formed successively from step 193, C10 begins to form structure C12 and then C12 folds to C11.

(11) At step 205, the population of C11 and C12 decreases, while ON state appears, suggesting a transition between these states. By forming a small hairpin b, most of the riboswitches fold through C11 into ON state with a free energy of $-73.07 \text{ kcal} \cdot \text{mol}^{-1}$.

At step 212, about 77% of the population is in ON state and about 11% of the population is in structure C12. According to Boltzmann distribution, the population of ON state ($-73.07 \text{ kcal} \cdot \text{mol}^{-1}$) and C12 state ($-67.96 \text{ kcal} \cdot \text{mol}^{-1}$) at equilibrium should be around 99.97% and 0.03%, respectively. Thus, the two states do not reach equilibrium at the end of the transcription. However, both states can induce downstream gene expression because of the structured splice site. Therefore, through a series of intermediate states, the TPP riboswitch finally folds into the genetic ON state to promote downstream gene expression without the ligand, in agreement with the previous study [23].

Our results suggested ON state of the TPP riboswitch was characterized by a four-way junction with a structured 5' splice site within helix P0 and hairpin b. As cotranscriptional folding follows a sequence process, the nucleotides of P0 helix in ON state (step 184) are transcribed prior to that of helix P1 (step 212) in the aptamer structure (OFF state). In addition, ON state is the most stable state in the absence of TPP ($\Delta G_{\text{ON}} = -73.07 \text{ kcal} \cdot \text{mol}^{-1}$, $\Delta G_{\text{OFF}} = -71.67 \text{ kcal} \cdot \text{mol}^{-1}$). Thus, even with TPP at a saturating level, the riboswitch still first folds to ON state, instead of the binding

pocket structure OFF, which is formed near the end of the transcription. For this riboswitch, whether the ligand is present or not, the main cotranscriptional folding pathway is the same. Unlike other riboswitches [10,33,46,47], sequential folding during the transcription context is not a prerequisite for the TPP riboswitch to efficiently perform gene regulation. Thus, ligand-induced conformational switching occurs after the full-length chain is transcribed.

The presence of TPP may aid flipping of the riboswitch

The cotranscriptional folding behavior of the TPP riboswitch indicated that whether TPP was present or not, the riboswitch predominantly folded to ON state, which was thermodynamically favored without TPP. When the ligand binds to the riboswitch, it will undergo conformational rearrangements to form the ligand-bound OFF state that prevents gene expression. Using the experimental measured values of the association and disassociation rates ($k_{\text{on}} = 8 \times 10^4 \text{ M}^{-1} \cdot \text{s}^{-1}$ and $k_{\text{off}} = 5 \times 10^{-2} \text{ s}^{-1}$) [26], the stabilized free energy caused by ligand binding under different TPP concentrations can be calculated as $\Delta G_{\text{binding}} = k_B T \ln(k_{\text{on}}[\text{L}]/k_{\text{off}}) \text{ kcal} \cdot \text{mol}^{-1}$ [10,33,35,48], where [L] is the TPP concentration. When the ligand is present at a high level, ON state formed first will equilibrate into the ligand-bound OFF state (OFF^b). However, the transition from ON state to OFF state requires

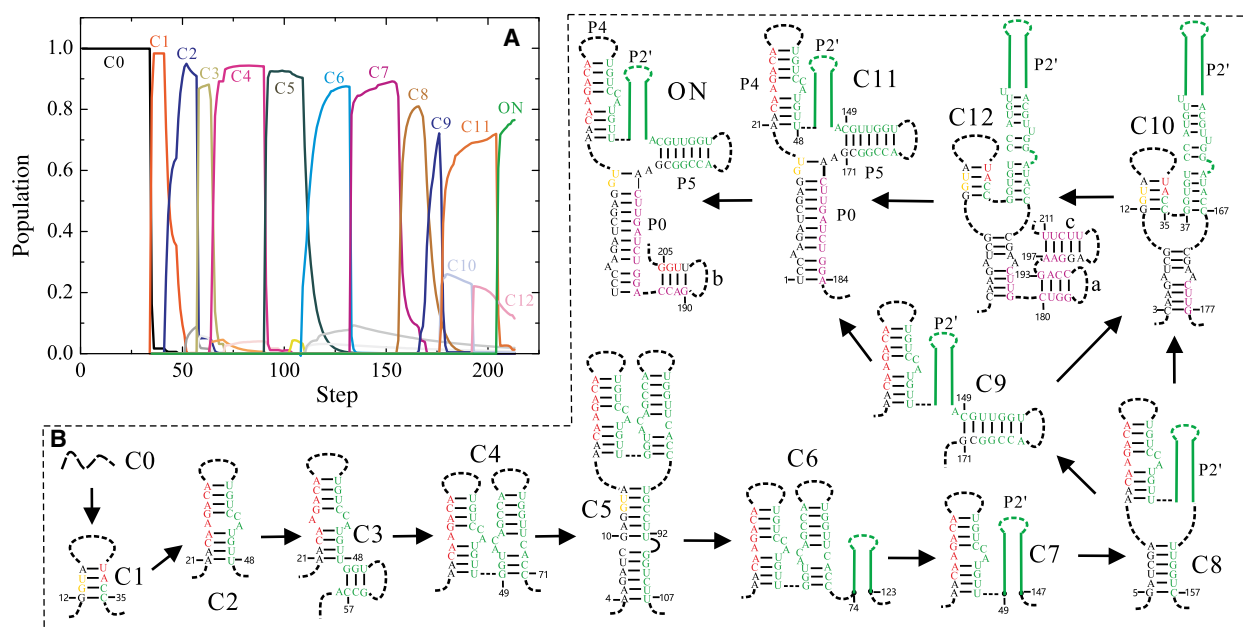


Fig. 3. Population kinetics (A) and the pathways (B) during the transcription of the TPP riboswitch at an elongation rate of $50 \text{ nt} \cdot \text{s}^{-1}$ without TPP. In (B), nucleotides are colored according to Fig. 1. The solid and dash lines denote paired regions and the loop or the connection regions, respectively.

undergoing enormous conformational rearrangements, as it needs to break helices P0, P4, and P5 and hairpin b to form helices P1, P2, and P3. Great conformational rearrangements, especially unfolding long helix regions, usually yield a very slow transition rate. The questions that what the transition pathway between the two functional states is and whether the transition rate is fast enough for this riboswitch to function remain unclear.

Helices P0, P4, and P5 of ON state shared nucleotides with helices P1, P2, and P3 of OFF state. Transition from ON state to OFF state should be extremely slow through completely unfolding of helices P0, P4, and P5 and hairpin b, followed by refolding of helices P1, P2, and P3 (with rate below 10^{-16} s^{-1}). As these helices overlap with each other, the most probable transition pathway is the tunneling pathway after unfolding hairpin b (Fig. 4) [39]. Thus, the ON–OFF transition can be divided into three main steps. Step (1): ON state transits to C11 by unfolding of hairpin b with a rate of about 200.637 s^{-1} . Step (2) involves two pathways: $\text{C11} \rightarrow \text{I1} \rightarrow \text{I2}$ and $\text{C11} \rightarrow \text{C9} \rightarrow \text{I2}$, which differ only in the order of formation of helices P2 and P3. As the bottom part of helix P2 overlaps with helices P4 and P5, the main way for the transition to form helix P2 ($\text{C11} \rightarrow \text{I1}$ and $\text{C9} \rightarrow \text{I2}$) is the tunneling pathway [39], where unfolding one stack in each helix

(P4 and P5) is followed by forming one stack in P2, returning a rate of 0.072 s^{-1} . To form helix P3 ($\text{I1} \rightarrow \text{I2}$ and $\text{C11} \rightarrow \text{C9}$) through the tunneling pathway, where unfolding one stack in P0 is followed by closing one stack in helix P3, yield a rate of 0.054 s^{-1} . Step (3) is the formation of the nonlocal helix P1 through the zipping pathway with a rate of 0.54 s^{-1} .

Although the tunneling pathway has lower transition barrier compared to the unfolding–refolding pathway [39], the transition from ON state to the aptamer structure through the tunneling pathway is still very slow due to the large conformational rearrangements. The overall transition rate from ON state to OFF state can be estimated to be around 10^{-5} s^{-1} , yielding a time much longer than the mean splicing time [36]. Since the ligand concentration in cells is much higher than that of mRNA, the transition rate from OFF state to OFF^b state can be assumed to be the effective binding rate $k_{\text{eff}} = k_{\text{on}}[\text{L}]$, and the reverse transition rate is the dissociation rate k_{off} . If binding of the ligand to the full-length riboswitch is described as a three-state transition model $\text{ON} \leftrightarrow \text{OFF} + \text{TPP} \leftrightarrow \text{OFF}^{\text{b}}$, flipping of the riboswitch will require about 83min when the ligand is present at $100 \mu\text{M}$ (Fig. 5). Even though OFF^b state is more stable than ON state ($\Delta G_{\text{binding}} = -3.14 \text{ kcal} \cdot \text{mol}^{-1}$) at this ligand concentration, the flipping time is much longer

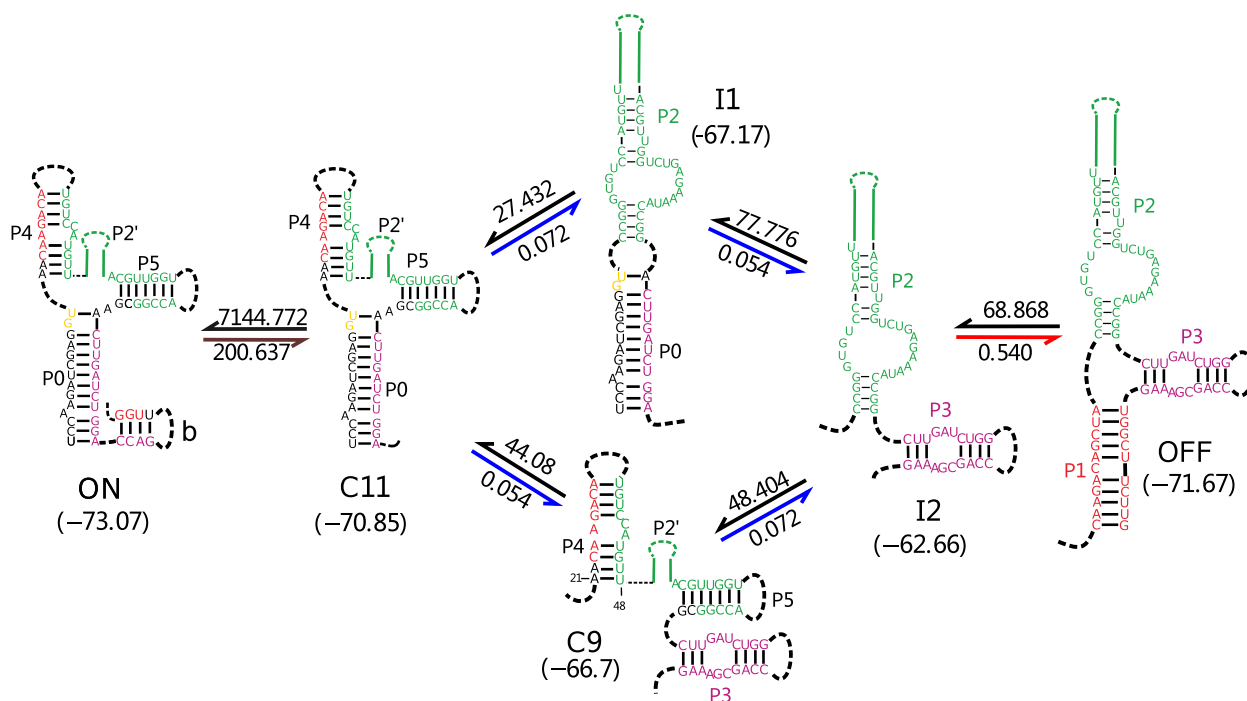


Fig. 4. The main structures on the transition pathway from ON state to OFF state. The free energies in parentheses and transition rates along the arrow are labeled in unit of $\text{kcal} \cdot \text{mol}^{-1}$ and s^{-1} , respectively. The transitions at steps (1), (2), and (3) are colored brown, blue, and red.

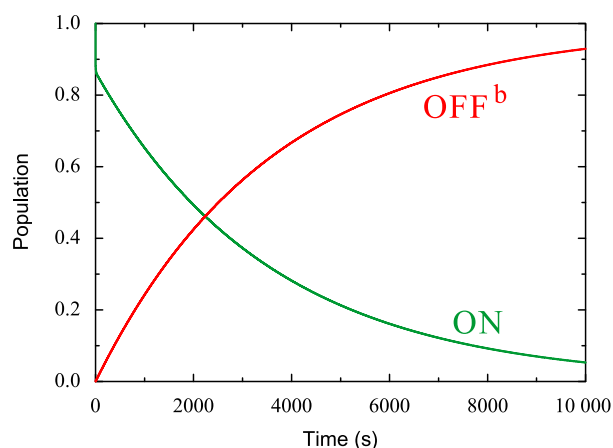


Fig. 5. Refolding from ON state to OFF^b state in the presence of 100 μ M TPP through the pathway: ON \leftrightarrow OFF + TPP \leftrightarrow OFF^b. The superscript 'b' denotes the ligand-bound OFF state.

than the half-time of splicing reaction [36]. Thus, the riboswitch is still in ON state when the splicing reaction begins. In this case, flipping of the switch is limited by the transition from ON state to OFF state, resulting in a constitutive-ON riboswitch behavior even with the ligand at a high level. To function as a TPP-sensed genetic switch, the RNA must fold to OFF^b state at least before the splicing reaction is ended. It means that TPP may interact with the riboswitch before OFF state is fully formed, thereby aiding the rest conformational rearrangements. Considering the crystal structure of OFF^b state, the transition pathway, and the free energies in Fig. 4, it can be speculated that helix P2 or helix P3 is likely to act as a platform for ligand to interact with the riboswitch and then lowers the barrier of the rest transition, which has been found in other riboswitches [49,50].

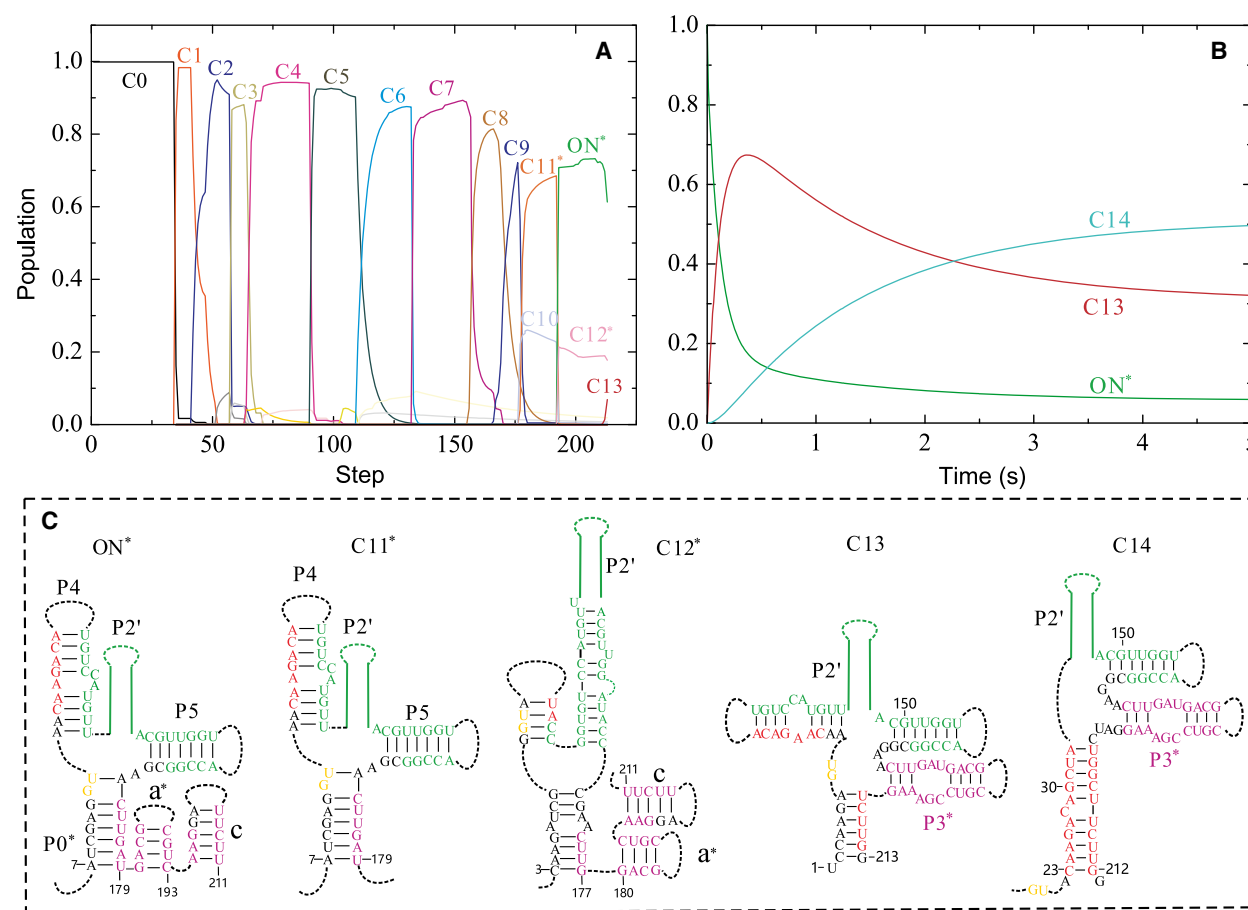


Fig. 6. Population kinetics of the mutated TPP riboswitch at 50 nt·s⁻¹ (A) and the refolding population kinetics where ON* state is the initial state (B). The main structures formed in (A) and (B) are shown in (C) and Fig. 3C. The upper * denotes the helices (P0*, P3*, and a*) and states (C11*, C12*, and ON*) formed by the mutated RNA construct with secondary structure that slightly differ from that of the wild-type.

Helix P0 is crucial for the TPP riboswitch to perform its function

The mutations in the P3 stem (Fig. 1) were found to be able to largely repress *NMTI* gene expression even when the ligand is absent [23]. This mutated TPP riboswitch exhibits only partial restoration of thiamine-dependent gene control, but at a level of expression that is far below the wild-type. To find the reason and explore the effect of the mutations on riboswitch activities, we studied the cotranscriptional folding pathway of the mutated RNA construct in the absence of TPP at a transcription rate of $50 \text{ nt}\cdot\text{s}^{-1}$ (Fig. 6A). The folding behavior is similar to that of the wild-type riboswitch before the 184th step. For the wild-type, helix P0 is fully formed at this step and state C9 transits to state C11 (Fig. 3), while most of the mutated RNAs transit from states C9 to C11*. Subsequently, state C11* folds to ON* state with a short nonlocal helix region. At the same time, about 22% of the mutated riboswitches fold from states C9 to C10 and then to C12*. Compared to the wild-type, the mutated helix P0* is shorter. Therefore, from step 193, states C11* and C10 can transit to states ON* and C12*, respectively, by forming the two hairpins a* and c. From about step 211, population of both states decreases, while another state C13 appears, suggesting a transition between these states. At the end of transcription, about 62% of the riboswitches adopt ON* state, while C12* and C13 occupy about 18% and 6% of the population, respectively.

For the mutated TPP riboswitch, OFF* state is the most stable structure ($\Delta G_{\text{OFF}^*} = -71.22 \text{ kcal}\cdot\text{mol}^{-1}$, $\Delta G_{\text{ON}^*} = -68.99 \text{ kcal}\cdot\text{mol}^{-1}$). Similar to the wild-type riboswitch, most riboswitches populated at the states (ON* and C12*) with a structured splice site region at the end of the transcription. Although the mutations destabilize helix P0, the rate of the transition from ON* state to OFF* state is still slow (around 10^{-4} s^{-1}). This implies that the mutated construct may also act as an on switch as the wild-type. However, ON* state equilibrates into structures C13 and C14 within 0.5 s (Fig 5B), which are more stable than ON* state ($\Delta G_{\text{C13}} = -70.42 \text{ kcal}\cdot\text{mol}^{-1}$, $\Delta G_{\text{C14}} = -70.85 \text{ kcal}\cdot\text{mol}^{-1}$). This result suggests that the riboswitches can significantly transit from ON* state to C13 and C14 before the splice reaction is initiated. Unlike ON* state, these two states have the unstructured splice site (Fig. 6C). Hence, even the transition between ON* and OFF* is still slow, the mutated riboswitch construct prevents gene expression in the absence of TPP in cells, consistent with experimental observation [23]. In other word, the mutated RNAs

will function as a ‘constitutive-OFF’ switch. As the mutations destabilize helix P0, the results suggest that the potential to fully form helix P0 is crucial for the TPP riboswitch to perform its switch function, which has been observed for S_{MK} riboswitch [3,35,51].

Conclusions

Riboswitches, as functional mRNAs, can adopt alternative folding conformations to control downstream gene expression in response to the ligand concentration. They can operate under different regulation regimes (kinetic versus thermodynamic). Most kinetically controlled riboswitches, such as pbuE [10,52], metF, and FMN riboswitches [12,33], usually exert regulatory control of transcription. For these riboswitches, ligand binding is not able to offset the free energy difference between ON and OFF states even at saturating concentrations. Once one functional state is formed, the switch cannot be flipped anymore. As the time window allowed for ligand binding to these riboswitches is controlled by the transcription process (transcription pausing and elongation rates), efficiencies of gene expression depend on the intracellular transcription context, instead of stabilities of the two functional structures. On the contrary, most thermodynamically controlled riboswitches, like *add* adenine and S_{MK} riboswitches [34,35], are reversible switches. Their regulation is in a ligand-dependent manner and not sensitive to the transcription context. These riboswitches can fluctuate readily between two functional states due to similar free energies and this fluctuation can occur multiple times before mRNA is degraded.

The regulatory behaviors of the TPP riboswitch within *NMTI* gene are coincident with a thermodynamic model. This riboswitch predominantly folds into the ON state, while the aptamer structure is not observed during the transcription. The predicted main cotranscriptional folding pathway of the TPP riboswitch is similar to that from Kinofold [53]. Its function is not sensitive to the transcription process, and ligand binding occurs post-transcriptionally. The mutated RNA construct largely loses the ability to perform genetic switch function, as a result of destabilization of the ON state.

The TPP riboswitch and S_{MK} riboswitch exhibit common features as other thermodynamic riboswitches. Besides, since both riboswitches only have a simple architecture that utilizes a single domain for ligand binding and gene regulation [35], they share some unique characters in regulatory behaviors. Their aptamer structures are not formed during the transcription. The potential to fully form nonlocal

helix P0 is crucial for both riboswitches to perform gene regulation [23,35,51]. But for the TPP riboswitch, ON–OFF transition needs to undergo large conformation rearrangements. Consequently, the transition between ON state and OFF state through helix exchanging pathways still returns a slower transition rate than the typical splice rate. To function as a switch, the ligand may interact with the riboswitch before full formation of the OFF state and, then, decrease the rest transition energy barrier. Further experimental and theoretical studies are required to fully understand the regulation of this riboswitch in cells.

Acknowledgements

This work has been supported by the National Natural Science Foundation of China (31600592).

Author contributions

SG wrote and designed the paper. All authors have contributed in the improving of the manuscript, and read and approved its final version.

References

- Serganov A and Nudler E (2013) A Decade of Riboswitches. *Cell* **152**, 17–24.
- Reining A, Nozinovic S, Schlepeckow K, Buhr F, Fürtig B and Schwalbe H (2013) Three-state mechanism couples ligand and temperature sensing in riboswitches. *Nature* **499**, 355–359.
- Lin J-C and Thirumalai D (2013) Kinetics of allosteric transitions in S-adenosylmethionine riboswitch are accurately predicted from the folding landscape. *J Am Chem Soc* **135**, 16641–16650.
- Perdrizet GA II, Artsimovitch I, Furman R, Sosnick TR and Pan T (2012) Transcriptional pausing coordinates folding of the aptamer domain and the expression platform of a riboswitch. *Proc Natl Acad Sci USA* **109**, 3323–3328.
- Mandal M, Boese B, Barrick JE, Winkler WC and Breaker RR (2003) Riboswitches control fundamental biochemical pathways in *Bacillus subtilis* and other bacteria. *Cell* **113**, 577–586.
- Winkler WC, Nahvi A, Sudarsan N, Barrick JE and Breaker RR (2003) An mRNA structure that controls gene expression by binding S-adenosylmethionine. *Nat Struct Biol* **10**, 701–707.
- Mandal M and Breaker RR (2004) Adenine riboswitches and gene activation by disruption of a transcription terminator. *Nat Struct Mol Biol* **11**, 29–35.
- Winkler WC and Breaker RR (2003) Genetic control by metabolite-binding riboswitches. *ChemBioChem* **4**, 1024–1032.
- Vitreschak AG, Rodionov DA, Mironov AA and Gelfand MS (2004) Riboswitches: the oldest mechanism for the regulation of gene expression? *TRENDS Genet* **20**, 44–50.
- Gong S, Wang Y and Zhang W (2015) Kinetic regulation mechanism of pbuE riboswitch. *J Chem Phys* **142**, 15103.
- Kim JN and Breaker RR (2008) Purine sensing by riboswitches. *Biol Cell* **100**, 1–11.
- Wickiser JK, Winkler WC, Breaker RR and Crothers DM (2005) The speed of RNA transcription and metabolite binding kinetics operate an FMN riboswitch. *Mol Cell* **18**, 49–60.
- Gong Z, Zhao Y, Chen C and Xiao Y (2011) Role of ligand binding in structural organization of add A-riboswitch Aptamer: a molecular dynamics simulation. *J Biomol Struct Dyn* **29**, 403–416.
- Miranda-Ríos J (2007) The THI-box riboswitch, or how RNA binds thiamin pyrophosphate. *Structure* **15**, 259–265.
- Kubodera T, Watanabe M, Yoshiuchi K, Yamashita N, Nishimura A, Nakai S, Gomi K and Hanamoto H (2003) Thiamine-regulated gene expression of *Aspergillus oryzae* thiA requires splicing of the intron containing a riboswitch-like domain in the 5'-UTR. *FEBS Lett* **555**, 516–520.
- Mukherjee S, Retwitzer MD, Barash D and Sengupta S (2018) Phylogenomic and comparative analysis of the distribution and regulatory patterns of TPP riboswitches in fungi. *Sci Rep* **8**, 1–13.
- Lünse CE, Scott FJ, Suckling CJ and Mayer G (2014) Novel TPP-riboswitch activators bypass metabolic enzyme dependency. *Front Chem* **2**, 1–8.
- Guedich S, Puffer-Enders B, Baltzinger M, Hoffmann G, Da Veiga C, Jossinet F, Thore S, Bec G, Ennifar E, Burnouf D *et al.* (2016) Quantitative and predictive model of kinetic regulation by *E. coli* TPP riboswitches. *RNA Biol* **13**, 373–390.
- Li S and Breaker RR (2013) Eukaryotic TPP riboswitch regulation of alternative splicing involving long-distance base pairing. *Nucleic Acids Res* **41**, 3022–3031.
- Rentmeister A, Mayer G, Kuhn N and Famulok M (2007) Conformational changes in the expression domain of the *Escherichia coli* thiM riboswitch. *Nucleic Acids Res* **35**, 3713–3722.
- Miranda-Ríos J, Navarro M and Soberon M (2001) A conserved RNA structure (thi box) is involved in regulation of thiamin biosynthetic gene expression in bacteria. *Proc Natl Acad Sci USA* **98**, 9736–9741.
- Croft MT, Moulin M, Webb ME and Smith AG (2007) Thiamine biosynthesis in algae is regulated by

- riboswitches. *Proc Natl Acad Sci USA* **104**, 20770–20775.
- 23 Cheah MT, Wachter A, Sudarsan N and Breaker RR (2007) Control of alternative RNA splicing and gene expression by eukaryotic riboswitches. *Nature* **447**, 497–500.
 - 24 Haller A, Altman RB, Soulière MF, Blanchard SC and Micura R (2013) Folding and ligand recognition of the TPP riboswitch aptamer at single-molecule resolution. *Proc Natl Acad Sci USA* **110**, 4188–4193.
 - 25 Cressina E, Chen L, Moulin M, Leeper FJ, Abell C and Smith AG (2011) Identification of novel ligands for thiamine pyrophosphate (TPP) riboswitches. *Biochem Soc Trans* **39**, 652–657.
 - 26 Anthony PC, Perez CF, García-García C and Block SM (2012) Folding energy landscape of the thiamine pyrophosphate riboswitch aptamer. *Proc Natl Acad Sci USA* **109**, 1485–1489.
 - 27 Kramer FR and Mills DR (1981) Secondary structure formation during RNA synthesis. *Nucleic Acids Res* **9**, 5109–5124.
 - 28 Lang K, Rieder R and Micura R (2007) Ligand-induced folding of the thiM TPP riboswitch investigated by a structure-based fluorescence spectroscopic approach. *Nucleic Acids Res* **35**, 5370–5378.
 - 29 Wachter A, Tunc-Ozdemir M, Grove BC, Green PJ, Shintani DK and Breaker RR (2007) Riboswitch control of gene expression in plants by splicing and alternative 3' end processing of mRNAs. *Plant Cell* **19**, 3437–3450.
 - 30 Yamauchi T, Miyoshi D, Kubodera T, Nishimura A, Nakai S and Sugimoto N (2005) Roles of Mg^{2+} in TPP-dependent riboswitch. *FEBS Lett* **579**, 2583–2588.
 - 31 Noeske J, Richter C, Stirnal E, Schwalbe H and Wöhnert J (2006) Phosphate-group recognition by the aptamer domain of the thiamine pyrophosphate sensing riboswitch. *ChemBioChem* **7**, 1451–1456.
 - 32 Zhao P, Zhang W and Chen S-J (2011) Cotranscriptional folding kinetics of ribonucleic acid secondary structures. *J Chem Phys* **135**, 245101.
 - 33 Gong S, Wang Y and Zhang W (2015) The regulation mechanism of yitJ and metF riboswitches. *J Chem Phys* **143**, 45103.
 - 34 Lemay J-F, Desnoyers G, Blouin S, Heppell B, Bastet L, St-Pierre P, Massé E and Lafontaine DA (2011) Comparative study between transcriptionally- and translationally-acting adenine riboswitches reveals key differences in riboswitch regulatory mechanisms. *PLoS Genet* **7**, e1001278.
 - 35 Gong S, Wang Y, Wang Z, Wang Y and Zhang W (2016) Reversible-switch mechanism of the SAM-III riboswitch. *J Phys Chem B* **120**, 12305–12311.
 - 36 Schmidt U, Basyuk E, Robert MC, Yoshida M, Villemain JP, Auboeuf D, Aitken S and Bertrand E (2011) Real-time imaging of cotranscriptional splicing reveals a kinetic model that reduces noise: implications for alternative splicing regulation. *J Cell Biol* **193**, 819–829.
 - 37 Xia T, SantaLucia J Jr, Burkard ME, Kierzek R, Schroeder SJ, Jiao X, Cox C and Turner DH (1998) Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick Base Pairs. *Biochemistry* **37**, 14719–14735.
 - 38 Mathews DH, Sabina J, Zuker M and Turner DH (1999) Expanded sequence dependence of thermodynamic parameters improves prediction of RNA secondary structure. *J Mol Biol* **288**, 911–940.
 - 39 Zhao P, Zhang W-B and Chen S-J (2010) Predicting secondary structural folding kinetics for nucleic acids. *Biophys J* **98**, 1617–1625.
 - 40 Zhang W and Chen S-J (2003) Master equation approach to finding the rate-limiting steps in biopolymer folding. *J Chem Phys* **118**, 3413–3420.
 - 41 Zhang W and Chen S-J (2006) Exploring the complex folding kinetics of RNA hairpins: II. Effect of sequence, length, and misfolded states. *Biophys J* **90**, 778–787.
 - 42 Chen J and Zhang W (2012) Kinetic analysis of the effects of target structure on siRNA efficiency. *J Chem Phys* **137**, 225102.
 - 43 Chen J, Gong S, Wang Y and Zhang W (2014) Kinetic partitioning mechanism of HDV ribozyme folding. *J Chem Phys* **140**, 25102.
 - 44 Wang Y, Wang Z, Liu T, Gong S and Zhang W (2018) Effects of flanking regions on HDV cotranscriptional folding kinetics. *RNA* **24**, 1229–1240.
 - 45 Geis M, Flamm C, Wolfinger MT, Tanzer A, Hofacker IL, Middendorf M, Mandl C, Stadler PF and Thurner C (2008) Folding kinetics of large RNAs. *J Mol Biol* **379**, 160–173.
 - 46 Heppell B and Lafontaine DA (2008) Folding of the SAM Aptamer is determined by the formation of a K-turn-dependent Pseudoknot. *Biochemistry* **47**, 1490–1499.
 - 47 Mandal M and Breaker RR (2004) Gene regulation by riboswitches. *Nat Rev Mol Cell Biol* **5**, 451–463.
 - 48 Gong S, Wang Y, Wang Z, Wang Y and Zhang W (2018) Genetic regulation mechanism of the yjdF riboswitch. *J Theor Biol* **439**, 152–159.
 - 49 Huang W, Kim J, Jha S and Aboul-ela F (2013) The impact of a ligand binding on strand migration in the SAM-I riboswitch. *PLoS Comput Biol* **9**, e1003069.
 - 50 Whitford PC, Schug A, Saunders J, Hennelly SP, Onuchic JN and Sanbonmatsu KY (2009) Nonlocal helix formation is key to understanding S-adenosylmethionine-I riboswitch function. *Biophys J* **96**, L07–L09.
 - 51 Lu C, Smith AM, Ding F, Chowdhury A, Henkin TM and Ke A (2011) Variable sequences outside the SAM-

- binding core critically influence the conformational dynamics of the SAM-III/SMK box riboswitch. *J Mol Biol* **409**, 786–799.
- 52 Neupane K, Yu H, Foster DAN, Wang F and Woodside MT (2011) Single-molecule force spectroscopy of the add adenine riboswitch relates folding to regulatory mechanism. *Nucleic Acids Res* **39**, 7677–7687.
- 53 Xayaphoummine A, Bucher T and Isambert H (2005) Kinefold web server for RNA/DNA folding path and structure prediction including pseudoknots and knots. *Nucleic Acids Res* **33**, 605–610.