RESEARCH ARTICLE | JANUARY 07 2015

Kinetic regulation mechanism of pbuE riboswitch

Sha Gong; Yujie Wang; Wenbing Zhang

J. Chem. Phys. 142, 015103 (2015)
https://doi.org/10.1063/1.4905214

Kinetic regulation mechanism of pbuE riboswitch

Sha Gong, Yujie Wang, and Wenbing Zhang[a)]
Department of Physics, Wuhan University, Wuhan, Hubei 430072, People's Republic of China

(Received 17 September 2014; accepted 17 December 2014; published online 7 January 2015)

Riboswitches are RNA residue segments located in untranslated regions of messenger RNAs. These folded segments directly bind ligands through shape complementarity and specific interactions in cells and alter the expression of genes at the transcriptional or translational level through conformation change. Using the recently developed systematic helix-based computational method to predict the cotranscription folding kinetics, we theoretically studied the cotranscription folding behavior of the *Bacillus subtilis* pbuE riboswitch in the absence and presence of the ligand. The ligand concentration, the transcription speed, and the transcription pausing are incorporated into the method. The results are in good agreement with the experimental results. We find that there are no obvious misfolded structures formed during the transcription and the formation of the ligand bound state is rate-limited by the association of the ligand and the RNA. For this kinetically driven riboswitch, the ligand concentration, the transcription speed, and the transcription pausing are coupled to perform regulatory activity. © *2015 AIP Publishing LLC.* [http://dx.doi.org/10.1063/1.4905214]

I. INTRODUCTION

Riboswitches are RNA residue segments located in untranslated regions of messenger RNAs that directly bind ligands through shape complementarity and specific interactions in cells and alter the expression of genes through conformation changes.[1–5] Riboswitches can recognize a wide variety of metabolites such as guanine,[6,7] glycine,[8,9] S-adenosylmethionine (SAM),[10,11] thiamine pyrophosphate (TPP),[12] lysine,[13] flavin mononucleotide (FMN),[14] coenzyme B_{12},[1] and adenine.[15,16] Most riboswitches are composed of two modular domains: an aptamer and an expression platform. The aptamer is the most conserved domain and is responsible for specifically recognizing the target metabolite. The expression platform, varying widely in sequence and structure, controls gene expression at level of transcription,[17,18] translation, and RNA splicing[12,19–22] by altering the mRNA structure.

Purine riboswitch aptamers form a three-way RNA junction from helix P1 and hairpins P2 and P3. The aptamers specifically bind the ligand through hydrogen bond interactions within the three-way junction.[23,24] Among the simplest riboswitches, the purine riboswitches display remarkable ligand selectivity and carry out markedly different functions despite the structural similarity of their aptamers. For the pbuE adenine riboswitch, in the absence of adenine, part of the aptamer region is involved in the formation of a terminator stem with the expression platform,[15,25] which results in transcription termination. The ligand binding activates gene expression through forming an antiterminator.[15,26] Previous studies have investigated many features of the purine riboswitch such as ligand specicity[27,28] and its structural basis,[28,29] folding landscape,[30,31] the rates and energies for ligand binding and dissociation,[32] the kinetics of loop-loop interaction,[33–35] and induced fit.[36,37] Most of them considered the folding of this aptamer. But the post-transcriptional full-length pbuE riboswitch was found to be fixed in the ligand-incompetent structure even when adenine is at saturation level,[32,35,38] which suggested that the regulatory activity should be in the transcription process.

During transcription elongation, because the upstream RNA section folds first, this will influence the folding pathway of the downstream RNA section. It has been proven that the RNA folding pathways and kinetic traps during transcription are different from those of during renaturation folding, and the transcription process dictates the RNA folding pathways and kinetic traps.[39–41] The transcription process was shown to have an important role for the regulatory activity of a FMN responsive riboswitch from *B. subtilis*.[42] For this kinetically driven riboswitch, besides the rates of ligand binding and riboswitch transcription affecting the regulation, additional factors such as transcriptional pause sites, which provide more time for the ligand to bind before the genetic decision is made, were also observed. Recently, an optical-trapping assay was developed to monitor cotranscription folding with force during the individual transcript of the pbuE riboswitch.[43] In this study, we predict the cotranscription folding kinetics in the absence/presence of the ligand by using the cotranscription folding theory[41] in which the elongation process is divided into many steps, each corresponding to the elongation of the chain by one nucleotide. The factors, which could affect the regulatory activity of the riboswitch, such as the ligand concentration, the transcription speed, and the transcription pausing are incorporated into the method.

II. THEORY AND METHOD

A. Master equation

For an ensemble of conformational states, the population $p_i(t)$ for each state i at time t can be described by the

a)Author to whom correspondence should be addressed. Electronic mail: wbzhang@whu.edu.cn

following equation (master equation): $dp_i(t)/dt = \sum_j^{\Omega}[k_{j\to i}p_j(t) - k_{i\to j}p_i(t)]$, where $k_{i\to j}$ and $k_{j\to i}$ are the rates of the respective transitions and Ω is the total number of conformations. The above equation can be written as the matrix form: $d\mathbf{p}(t)/dt = \mathbf{M}\bullet\mathbf{p}(t)$, where $\mathbf{p}(t)$ is the vector for the population distribution, $\mathbf{M}$ is the $\Omega\times\Omega$ rate matrix with the elements defined by $M_{ij} = k_{i\to j}(i\neq j)$, and $M_{ii} = -\sum_{j\neq i} k_{i\to j}$. Solving the equation, the probability kinetics for a given initial condition can be written as

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

Where λ_m and $\mathbf{n}_m$ are the m-th eigenvalue and eigenvector of rate matrix $\mathbf{M}$, respectively, and C_m is the coefficient that is dependent on the initial condition.

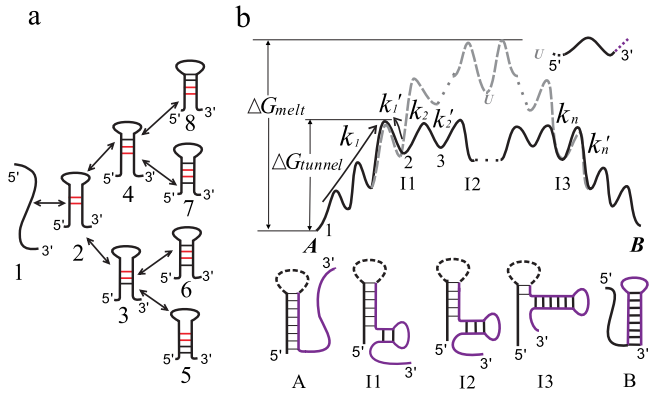

FIG. 1. (a) The formation of a helix along the zipping pathways. The first closing stack denoted by the red colour. (b) The free energy landscape for the transition between helix A and B. Solid line: the tunneling pathway. Dotted line: completely unfolding helix A followed by refolding to B.

B. Helix based kinetic model

In our model, the rates for formation (k_+) and disruption (k_-) of a base stack can be calculated as[44] $k_+ = k_0 e^{-\Delta G_+/k_B T}$ and $k_- = k_0 e^{-\Delta G_-/k_B T}$, where ΔG_+ and ΔG_- are the energy barriers for formation and disruption of a stack, respectively, k_0 is a prefactor, k_B is the Boltzmann constant, and T is the temperature. The barrier for formation of the transition state is entropic: $\Delta G_+ = T\Delta S$. The barrier for disruption of the transition state is enthalpic: $\Delta G_- = \Delta H$. Thus, the rates for formation and disruption of a stack are $k_+ = k_0 e^{-\Delta S_{\text{stack}}/k_B}$ and $k_- = k_0 e^{-\Delta H_{\text{stack}}/k_B T}$. The rates for formation and disruption of a loop-closing stack (and the loop) are $k_+^{\text{loop}} = k_0 e^{-(\Delta S_{\text{stack}}+\Delta S_{\text{loop}})/k_B}$ and $k_-^{\text{loop}} = k_0 e^{-\Delta H_{\text{stack}}/k_B T}$, where $-\Delta S_{\text{stack}}$ and $-\Delta H_{\text{stack}}$ are the entropy and enthalpy changes upon formation of the stack, respectively, and $-\Delta S_{\text{loop}}$ is the entropy change of the loop. The prefactor k_0 has been fitted from the experimental data and is equal to $6.6\times10^{12}\text{s}^{-1}$ for the formation/disruption of an AU base pair and $6.6\times10^{13}\text{s}^{-1}$ for a GC base pair.[41,44–47]

As the rate for formation of a base stack is usually larger than that of disrupting the stack except the loop-closing stack under the folding condition, once the first few stacks in a helix are closed and stabilized, zipping of the subsequent stacks in the helix would be fast and it would quickly slip into the fully folded helix.[44] It is proper to use the helices as building blocks for the study of the overall (slower) folding kinetics. The conformational space for the kinetics calculation can be drastically reduced by using helices as the building blocks. In this model, different structures are constructed according to the helices as the building blocks and the kinetic move are the addition or deletion of a helix or the exchange between two helices.

C. Rate of formation and disruption of a helix

If two conformations differ only in one helix, the transition between them would be the formation and disruption of the helix. The most probable pathway for the helix formation is the zipping pathway (Fig. 1(a)).[44] Since the rate for forming the first loop-closing stack is much smaller than that for adding a new stack to the existing stack, the formation of the helix is rate-limited by the formation of the first stack (from state 1 to 2). Considering that a fraction of population which "flows" from state 1 to state 2 would bounce back, the zipping rate k_f along the $1\to2\to3$ pathway in Fig. 1(a) can be calculated as

$$k_f = k_{1\to2}K_1(1 - K_2'K_1'\sum_{n=0}^{\infty}(K_2'K_1)^n) = k_{1\to2}K_1(1 - K_2'K_1'\frac{1}{1-K_2'K_1}). \quad (2)$$

Where $k_{i\to j}$ denotes the rate for the transition from state i to state j, K_1 and K_1' are the forward (state $2\to3$) and reverse (state $2\to1$) probabilities for state 2, and K_2 and K_2' are the forward (state $3\to5$ and $3\to6$) and reverse (state $3\to2$) probabilities for state 3

$$K_1 = \frac{k_{2\to3}}{k_{2\to3}+k_{2\to1}+k_{2\to4}}, K_1' = \frac{k_{2\to1}}{k_{2\to3}+k_{2\to1}+k_{2\to4}},$$
$$K_2 = \frac{k_{3\to5}+k_{3\to6}}{k_{3\to2}+k_{3\to5}+k_{3\to6}}, K_2' = \frac{k_{3\to2}}{k_{3\to2}+k_{3\to5}+k_{3\to6}}.$$

For a given RNA molecule, the first base stack can be formed anywhere inside the helix. Therefore, the net rate k_F for the formation of a helix is the sum of the rates along all the pathways (Fig. 1(a)) with the different first (nucleation) base stacks. The rate for the disruption of the helix can be estimated from the equilibrium constant of the helix: $k_U = k_F e^{-\Delta G/k_B T}$, where ΔG is the folding free energy of the helix.

D. Rate of exchanging between two helices

If two helices A and B overlap with each other, they cannot coexist in the same structure. The conversion from helix A to helix B through complete unfolding of helix A followed by refolding to B is extremely slow due to the high energy barrier to disrupt all the base stacks in helix A. The tunneling pathway—where, at first, helix A is partially disrupted, and in each subsequent step, disruption of a base stack in A is accompanied by a concurrent formation of a base stack in B (see Fig. 1(b))—is much faster with a lower energy barrier. According to the free energy landscape, the tunneling pathway

for the helix A → helix B transition can be classified into three stages. Stage 1: from state A to I1, it is an uphill process on the free energy landscape; stage 2: from state I1 to I3, it is an oscillating process on the free energy landscape; stage 3: from state I3 to B, it is a downhill process on the free energy landscape. Stages 1 and 3 can be treated as two-state transitions without significant accumulation of intermediate states. At stage 2, all the intermediate states between the states I1 and I3 have high free energies and no significant population would aggregate for these states during the kinetic process. Using steady-state approximation for these intermediate states, the rates between states A and B for helix exchanging can be estimated as[44]

$$k_{A\to B} = \frac{\prod_i^n k_i}{\sum_{j=0}^{n-1}(\prod_{i=1}^{j} k_i' \prod_{m=j+2}^{n} k_m)}, k_{B\to A} = k_{A\to B} e^{-\Delta G_{AB}/k_B T}, \quad (3)$$

where k_n and k_n' are the rate constants for the process of formation (disruption) and disruption (formation) of a base stack in A and B, respectively.

E. Cotranscription folding kinetics

In this model,[41] releasing one nucleotide by the RNA polymerase (RNAP) to freely form secondary structures is treated as a transcriptional step. If the transcription speed of a RNA sequence is v nucleotides per second, the (real) time window for each step would be $1/v$ s, i.e., the polymerase spends $1/v$ s to synthesize a nucleotide. From time t, when an N-th nucleotide (nt) chain is (newly) transcribed to time $t+1/v$ and when the $(N+1)$-th nucleotide is (newly) transcribed, the N-th nt chain samples, the conforma tional space, and its population distribution are relaxed from $[P_1(N)_{\text{begin}}, P_2(N)_{\text{begin}}, \ldots, P_\Omega(N)_{\text{begin}}]$ to $[P_1(N)_{\text{end}}, P_2(N)_{\text{end}}, \ldots, P_\Omega(N)_{\text{end}}]$, where $P_i(N)_{\text{begin}}$ and $P_i(N)_{\text{end}}$ are the population of state i at the beginning and end of step N, respectively. Ω is the number of conformations for the N-th nt chain. This is defined as the N-th step. For each such step, the complete conformational ensemble of the N-th nt chain is first generated, and then the transition rates between the different structures (see Eqs. (2) and (3)) and the population kinetics from time t to $t+1/v$ (see Eq. (1)) are calculated.

The beginning population of the $(N+1)$-th step is inherited from the ending population of the N-th step. However, the RNA chain in the $(N+1)$-th step is one nucleotide longer than in the N-th step. According to the possible changes of the structures upon the elongation of the chain by one nucleotide, the structures are classified as four types (see Fig. 2). Type a and type b: The newly transcribed $(N+1)$-th nucleotide is added to the 3′ end of an open chain and the dangling tail of a helix, respectively. In these two categories, the new $(N+1)$-th nt chain can retain the same secondary structure as the N-th nt chain as the addition of the new nucleotide will not cause change of the existing structure. Type c: The newly transcribed nucleotide can pair with an upstream nucleotide to form an elongated helix by one base pair. Since the zipping of the new base pair (stack) (with time ~ (rate)$^{-1}$ ~ 10^{-6} s) is much faster than transcribing a nucleotide ($1/v$ from 2.5×10^{-3}s to 5×10^{-2}s),[41] these two structures can be treated as "directly inherited" and thus have the same population. Type d: All the new $(N+1)$-th nt structures that cannot be formed in an N-th nt chain and thus have zero population at the beginning of the $(N+1)$-th step. The population distribution at the beginning of step $N+1$ can be summarized by the equations below $P(N+1)_{\text{begin}} = P(N)_{\text{end}}$ for a, b, and c and $P(N+1)_{\text{begin}} = 0$ for d.

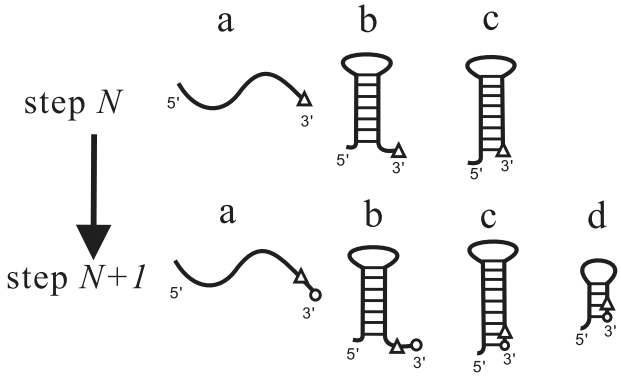

FIG. 2. Four types of relationships between N-th nt and $(N+1)$-th nt structures. The triangle and circle denote the last transcribed nucleotides at step N and $N+1$, respectively.

For consecutive steps, the folding results of the previous step turn into the initial condition of the next step. Applying this method from the first step to the end of transcription, we compute the folding kinetics for the RNA chain during transcription. To simulate transcriptional pausing at a specific site, we can assign a large number of effective time steps for the corresponding (paused) step.

F. Ligand binding kinetics

To simulate the cotranscription folding behavior of the riboswitch in the presence of adenine, we incorporated the ligand binding kinetics into the model. As the concentration of the ligand is much larger than that of the mRNA in the cell, the second order ligand binding kinetics, mRNA + ligand → the ligand bound state, can be approximated as a linear relation. The ligand bound states are added to the conformation space when adenine is present. The ligand free structures can transit to the corresponding ligand bound states with the effective binding rate $k_{\text{eff}} = k_{\text{on}}[\text{A}]$; the ligand bound states transit to the ligand free states with the rate k_{off}, and the ligand binding is able to stabilize the corresponding bound state by a free energy of $\Delta G_{\text{binding}} = k_B T \ln(k_{\text{on}}[\text{A}]/k_{\text{off}})$, where [A] is the concentration of the ligand, k_{on} is the association rate, and k_{off} is the dissociation rate. Except these two rates, all the other transition rates between the states in the conformation space are calculated with Eqs. (2) and (3). In our calculation, only the structures with helices P1, P2, and P3 are capable of ligand binding, and the values of k_{on} and k_{off} are taken from experimentally measured values as $k_{\text{on}} = 8\times10^4\,\text{M}^{-1}\text{s}^{-1}$ and $k_{\text{off}} = 0.15\,\text{s}^{-1}$.[48]

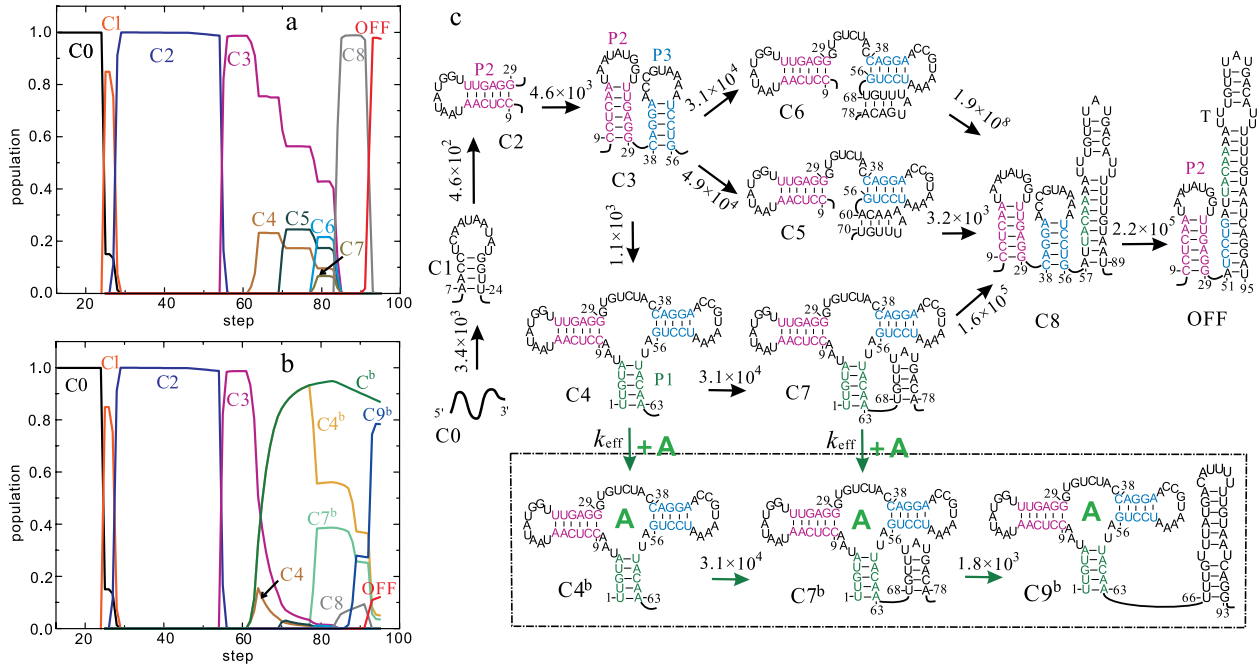

FIG. 3. The population kinetics of main states (C0–C9) formed during the transcription of the pbuE riboswitch without (a) and with (b) 300 μM adenine at a transcription rate of 20 nt/s, the transition pathway and the structure of the states (c). The superscript "b" denotes the corresponding state with adenine bound and C^b denotes the ensemble of all the bound states (C4^b, C7^b, and C9^b). In (c), the nucleotides within the P1, P2, and P3 paired region are colored differently: P1(green), P2(fuchsia), and P3(blue). T is the terminator hairpin. All the labeled transition rates (unit s^{-1}) along the arrow except k_{eff} are calculated with Eqs. (2) and (3). $k_{\text{eff}} = k_{\text{on}}[\text{A}]$ is the effective ligand binding rate, where k_{on} is taken from experimentally measured value and [A] is the concentration of the ligand.

III. RESULTS AND DISCUSSION

A. Cotranscription folding behavior of pbuE riboswitch in the absence/presence of ligand

Since the termination point occurs near the last U (the 95th nucleotide) within the U-stretch,[49–51] the cotranscription folding kinetics of the 95-th nt full-length pbuE riboswitch was studied with this method step by step. The cotranscription folding kinetics of the full-length pbuE riboswitch without adenine at a typical elongation rate of 20 nt/s are shown in Fig. 3(a). As the chain grows, the nascent RNA chain folds through a series of discrete intermediate states (from C0 to C9) to the OFF state, which consists of helix P2 and the terminator hairpin (the T hairpin). The free energies, which are calculated according to the Turner rule,[52,53] of the main intermediate states are listed in Table I. The transition rates between the states calculated with Eqs. (2) and (3) are also shown in Fig. 3. It shows that along the transcription, helix P2 is first formed as the first 29-th nt is transcribed and it will stay in this structure till the 55th nucleotide is transcribed. And then, as the 56th nucleotide is free to form structures, a branch structure C3 consisting of helices P2 and P3 is formed and it occupies most of the population till the 61st step. Helix P1 can be formed when the 62nd nucleotide is transcribed, and then part of the population of structure C3 will transit to the aptamer (structure C4), which consists of helices P1, P2, and P3. The free energy of C4 is higher than that of C3 structure ($\Delta G_{\text{C4}} = -11.69$ kcal/mol and $\Delta G_{\text{C3}} = -12.42$ kcal/mol), as helix P1 is unstable because it has only 4 stacks and closes a multi-branch loop. The transition rates between C3 and C4 are $k_{\text{C3}\to\text{C4}} = 1.1\times10^3$ s^{-1} and $k_{\text{C4}\to\text{C3}} = 3.4\times10^3$ s^{-1}, respectively. The two conformations can reach equilibrium in 1 and 2 steps and each occupies a proportion of population about 75% and 23%, respectively. At about step 70, structure C3 transits to structure C5 by adding a short helix. From step 75 to step 80, the upper part of the T hairpin can be formed without breaking helix P1, and structures C3 and C4 will transit to structures C6 and C7. Structures C3, C4, C5, C6, and C7 can get to equilibrium at step 80, and each occupies about 42.9%, 9.6%, 17.4%, 21.6%, and 6.7% of the population, respectively. From step 83, the T hairpin begins to invade helix P1, and this transition is of the tunneling pathway (replacing each base pair successively) with a very large rate (1.6×10^5 s^{-1}). Till step 85, helix P1 will be replaced by the T hairpin and almost all the population is occupied by structure C8. Further transcription to step 90, the T hairpin begins to invade helix P3 and when the RNAP moves to the termination point (at step 95), helix P3 will be fully replaced by the T hairpin, and the population of the OFF state with the T hairpin reaches about 98%. Considering the factors (RNAP, ribonucleoside triphosphate (NTP), etc.)

TABLE I. Free energies of the intermediate states formed during the transcription.

State	C0	C1	C2	C3	C4	C5	C6	C7	C8	C9	OFF
ΔG (kcal/mol)	0	−1.07	−7.67	−12.42	−11.69	−11.90	−12.19	−11.46	−22.98	−13.30	−29.33

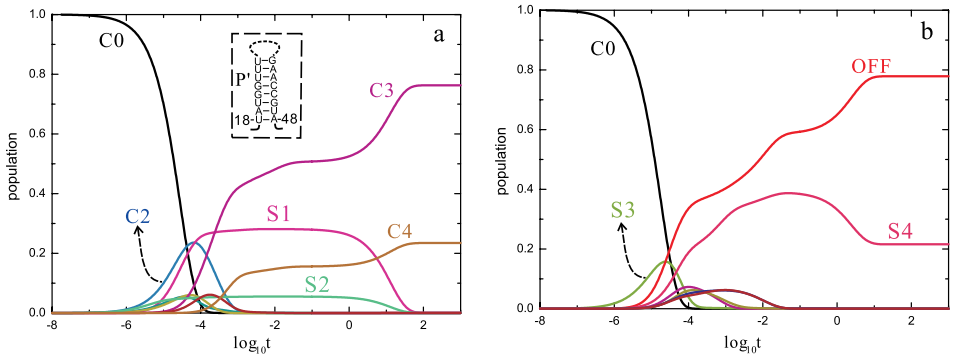

FIG. 4. The refolding population kinetics of the 64-th nt aptamer (a) and the 95th-nt full-length pbuE riboswitch (b). Structures S1 and S4 consist of helices P1 and P′, and helix P′ and the T hairpin, respectively. Structures S2 and S3 consist of helix P′ and the T hairpin, respectively.

which may affect the termination efficiency (TE),[54,55] the result agreed well with TE of 89% given by the gel-based assay.[43] The folding behavior of the riboswitch without adenine is consistent with the observation that the P2 helix and T helix are formed in turn.[43]

During the transcription, the aptamer which consists of helices P1, P2, and P3 is formed prior to the T hairpin, and as the transition rate from this structure to the OFF state is very large through a tunneling pathway, almost all the population will transit to the OFF state even with a transcription rate as high as 400 nt/s at the end of transcription. Thus, the aptamer cannot act as a trap to prevent the formation of the T hairpin without adenine. The free energy of the OFF state ($\Delta G_{\mathrm{OFF}} = -29.33$ kcal/mol) is much lower than that of the aptamer ($\Delta G_{\mathrm{C4}} = -11.69$ kcal/mol) and that of the ligand bound state, considering that the ligand binding would contribute a free energy as low as (4 ± 1) kcal/mol.[48] After the transcription, the OFF state would not transit to the ligand bound state even in the presence of adenine with a saturation concentration. Experimentally,[15,32] the OFF state of this riboswitch was found to hinder adenine binding even at a high ligand concentration. During the transcription, there were also no other misfolded states formed as a trap to prevent the formation of the T hairpin. Thus, the binding process must occur during the transcription before the formation of the OFF state.

It has been suggested that helix P1 only needed to form about two base pairs and then the aptamer could bind the ligand.[25] So at each step in our calculation with the ligand presence, each structure (ligand-competent state) that consists of helices P1 (at least having two base pairs), P2, and P3 could transit to the corresponding ligand bound state. The ligand binding kinetics are incorporated into the model by adding the ligand bound states to the conformation space and the forward and reverse rates of the transition from the ligand-competent structure to the corresponding ligand bound state are k_{eff} and k_{off}, respectively. The cotranscription folding process of pbuE riboswitch with 300 μM adenine at 20 nt/s is shown in Fig. 3(b), where the corresponding bound states are defined by $C4^b$, $C7^b$, and $C9^b$. It shows the same folding behavior as that in the absence of the ligand before the 62nd step, since that the aptamer begins to form and then transits to the ligand bound state $C4^b$ through structure C4 (Fig. 3(c), the green arrow). The population of $C4^b$ continues to increase and till the 79th nucleotide is transcribed, almost all the population is occupied by the ligand bound state at this saturation concentration. Similar to that in the absence of adenine, a part of structure $C4^b$ would transit to structure $C7^b$ by forming the upper part of the T hairpin and these two structures can attain equilibration. From step 85, the free energy of structure C8 is lower than that of the bound states, and the T hairpin begins to invade the aptamer. But the invasion in the presence of adenine is much slower than that without adenine. As the 90th nucleotide is transcribed, structure $C9^b$ consisting of the aptamer and a new hairpin other than the T hairpin is formed. This structure can obtain about 78% of the population at the end of transcription, while it has little population without the ligand. When the full-length riboswitch is synthesized, about 87% of the population is in the genetic on state.

During the transcription process, the aptamer structure (C4) can be formed before the OFF state, and adenine can bind to the aptamer structure and then the riboswitch would fold into the genetic on state. However, the full-length riboswitch refolds to the thermodynamically favored OFF state in less than 1 s while the aptamer structure does not emerge (see Fig. 4(b)). Therefore, the ligand binding must occur during the transcription process and the action of the riboswitch is sensitive to the cotranscription folding.

B. The effects of adenine concentration and transcription rate on the termination efficiency

The pbuE gene encodes a purine efflux pump, and adenine sensing by the riboswitch allows gene expression of this pump, presumably to reduce intracellular ligand concentrations.[28,56] The forming of the bound states can be modulated not only by the concentration of the ligand but also by the elongation rate and the proper transcription pausing, which can be regulated by RNAP, the concentration of NTP, proteins, and other cellar environments.[42,57,58] We will study the effect of adenine concentration and the transcription rate on the termination efficiency at this section.

Fig. 5 shows that in the presence of 30 μM adenine, the population of the OFF state would reach about 67% and 83% with the transcription rate 20 nt/s and 50 nt/s, respectively. At 100 μM ligand, the values are about 31% and 55%, respectively. It implies that excessive adenine would result in the expression of pbuE-encoded genes to pump out the ligand, and a higher concentration of the ligand is needed to trigger the riboswitch in genetic on state with a faster transcription

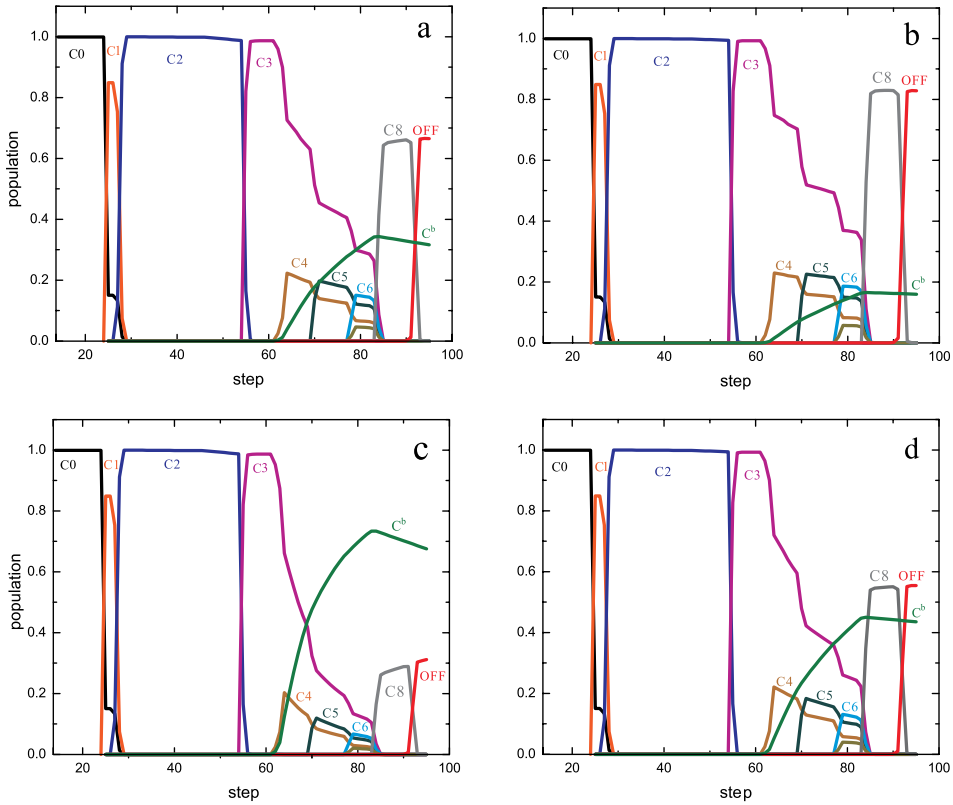

FIG. 5. Cotranscription folding process with different elongation rates in the presence of 30 μM ligand: (a) 20 nt/s, (b) 50 nt/s; 100 μM ligand: (c) 20 nt/s, (d) 50 nt/s.

rate. The slow transcription rate provides more time for ligand binding, leading to a high prevalence of the bound states, which has also been found in the FMN riboswitch.[42]

The aptamer can be formed when the 62nd nucleotide is transcribed and then the ligand can bind to the aptamer. Ligand binding can provide the ligand bound states a free energy of about −1.72 to −3.14 kcal/mol, if the concentration of adenine is about 30–300 μM, considering $\Delta G_{\text{binding}} = k_B T \ln(k_{\text{on}}[\text{A}]/k_{\text{off}})$. The bound state C4^{b} is the most stable structure before the (84 ± 1)th nucleotide is transcribed. And after that point, the free energy of C8, in which helix P1 is invaded by the T hairpin, is even lower than that of the bound states. The ligand binding is a kinetic process, so the amount of the ligand that could bind to the aptamer depends on the effective binding rate and the temporal window. C3 is the most stable structure and almost all the population is occupied by this structure before C4, to which the ligand binds, can be formed. The transition rate from C3 to C4 ($1.1 \times 10^3\ \text{s}^{-1}$) is much faster than k_{eff} (the transition rate from C4 to C4^{b}), which is $2.4\ \text{s}^{-1}$ and $8.0\ \text{s}^{-1}$ at 30 μM and 100 μM adenine, respectively. The forming of the ligand bound state is rate-limited by the association of the ligand and the RNA. Also the transition rate (about $3.1 \times 10^5 \text{s}^{-1}$) from structure C3 to C8 is much larger than k_{eff}. So if the ligand cannot bind to the aptamer in the temporal window $(84-62)/v$, where v is the transcription rate, C3 would transit to C8 and then to the OFF state. As the transition rate from the ligand-competent state to the corresponding ligand bound state is $k_{\text{on}}[\text{A}]$, the population that the bound states (C^{b}) can obtain within the time period $22/v$ is proportional to $A_{\text{equil}}(1 - \exp(-k_{\text{on}}[\text{A}] \times \frac{22}{v}))$, where A_{equil} is the equilibrium population of the bound states before step 84, which is about 74% and 90% when the ligand concentrations are 30 μM and 100 μM, respectively. Thus, the ligand concentration and transcription rate coupled to regulate the riboswitch on-off switch. When $k_{\text{on}}[\text{A}] < v/22$, the temporal window is shorter than the time needed for the ligand binding, then the population of the bound states is less than the equilibrium population. E.g., as shown in Fig. 5 at 30 μM adenine, the bound states can only reach about 31% and 16% when the transcription rate is 20 nt/s and 50 nt/s, respectively. When the ligand concentration is high enough as $k_{\text{on}}[\text{A}] \gg v/22$, most of the population would be in the on state (see Fig. 3).

C. The effect of transcription pausing on the activities of pbuE riboswitch

For kinetically driven riboswitches, the time needed for RNA polymerase to progress from the 3′ end of the aptamer to the intrinsic terminator is an important factor that influences the activities of the riboswitches. Previous studies suggested the actual recognition process of adenine was likely to be slow and transcriptional pausing might be required to "wait" for aptamer folding and ligand binding to avoid misfolded, nonresponsive structures.[15,59] As shown above, there are no obvious misfolded structures formed during the transcription and the formation of the bound states is rate-limited by the association of the ligand and the RNA. So transcription pausing, which often occurs at U-rich sites,[50,60] will provide extra time for ligand binding. For the pbuE riboswitch, there are two short series of uridine (U) residues, situated at 79-84 and 95-102, respectively. Considering the 11 and 12 nt enclosed by the RNAP from 3'terminal,[61,62] the possible pausing points situated at step 67 and 83, to which the

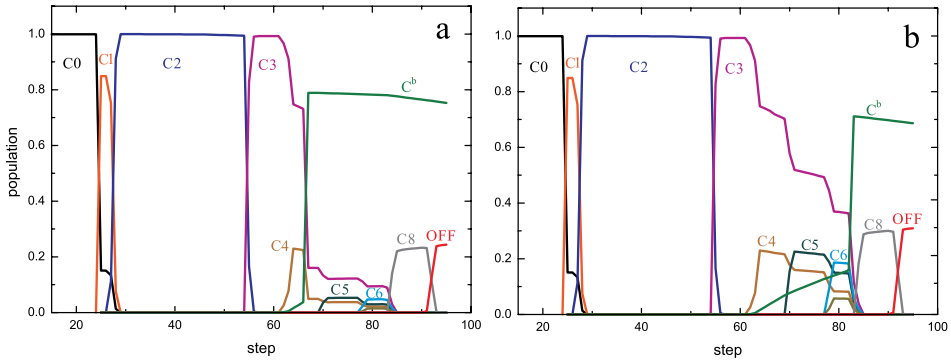

FIG. 6. Cotranscription folding with pausing (60 s) at step 67 (a) and 83 (b) in the presence of 30 μM adenine at a transcription rate of 50 nt/s.

nucleotides are free from the RNAP and are capable to form structures. The ligand bound state is the most stable structure from step 62 to step 83, so pausing anywhere between step 62 and 83 (see Fig. 6) would not make much difference. It is in accordance with the fact that the ligand must bind before step 84 to prevent the thermodynamically favored T hairpin formation.

To simply assess the effect of pausing for the function of the riboswitch, pausing was presumed to occur at the 67th step for a long time to ensure the equilibrium attained significantly. It shows when the transcription pauses about 60 s, the population of the OFF state would decrease from about 83% (see Fig. 5(b)) to 25% (see Fig. 6(a)) with 30 μM adenine at 50 nt/s, which implies that pausing would result in less ligand to trigger the riboswitch to the genetic on state. Assuming the pausing time τ, the population that the bound states can attain within the time period $22/v+\tau$ is proportional to $A_{\text{equil}}(1-\exp(-k_{\text{on}}[\text{A}]\times(\frac{22}{v}+\tau)))$. The effect of pausing is not so obvious for the function of this riboswitch when the concentration of adenine is high or the elongation rate is very slow. For instance, when the transcription rate is 20 nt/s and the ligand concentration is 300 μM, the binding equilibrium can be significantly attained within the binding time window without pausing. But for adenine at a relatively low concentration, especially at a fast elongation rate, the pausing time would affect the outcome of the riboswitch as it can hardly get to equilibrium without pausing. When the pausing time is long enough for the ligand bound states to reach equilibrium, the regulation mechanism of this riboswitch will present the characters of thermodynamic mechanism.

IV. CONCLUSION

Using a RNA cotranscription folding kinetics theory based on the creation/disruption and exchange of helices, we investigate the cotranscription folding behavior of the pbuE riboswitch at discrete steps. The riboswitch shows a kinetic regulation mechanism. The transcription rate, the ligand concentration, and the pausing are coupled to regulate the gene expression. Cotranscription folding of this riboswitch directly determines the fate of the transcription.

The full-length pbuE riboswitch is fixed at a ligand-incompetent structure and the refolding kinetics of the full-length riboswitch show that the aptamer is not formed, suggesting that the transcriptional context is critical for ligand binding activity. Although the aptamer is formed at step 62 during the transcription, it will be invaded by the T hairpin at the 84th step. The time period allowed for ligand binding is $22/v$, which is too short for the bound states to get a high proportion of the population when the ligand concentration is low. The two U-stretch sites, just corresponding to step 67 and 83 in our calculation, are possible to act as transcription pausing sites, which may provide extra time for the ligand binding to reduce the demand of adenine for the riboswitch to promote activity. Assuming the pausing time τ, the population that the bound states can attain is proportional to $A_{\text{equil}}(1-\exp(-k_{\text{on}}[\text{A}]\times(\frac{22}{v}+\tau)))$. For this kinetically driven riboswitch, the ligand concentration, the transcription speed, and the transcription pausing are coupled to perform regulatory activity. When the ligand concentration is high, the elongation rate is slow or transcription is pausing, most of the riboswitches can bind to the ligand before the OFF state is thermodynamically favored, and then this riboswitch will be trapped into the bound states and induce the downstream gene expression. When the pausing time is long enough for the bound states to reach equilibrium, the regulation of this riboswitch will present the characters of thermodynamic mechanism. Since pausing occurs usually because the RNA 3′ OH and the NTP substrate are failed to maintain alignment in the active site,[55] it depends on the transcription conditions such as NTP concentration, RNAP, proteins, and other factors and these factors also affect the transcription rate.[42,55,60] Therefore, downstream gene expression of the pbuE riboswitch is largely dependent on the cell environment. However, the current method has limitations. The free energy parameters at this model are of RNA at 1M NaCl solution condition and the effect of Mg^{2+} ions on these parameters is neglected. As Mg^{2+} can significantly stabilize the tertiary interactions,[63–67] it may alter the folding pathways. Future development of the model would incorporate the Mg^{2+} effects into the model.

ACKNOWLEDGMENTS

This work was partly supported by the National Natural Science Foundation of China under Grant No. 31270761 and the Ph.D. Programs Foundation of Ministry of Education of China under Grant No. 20110141110009 (to W. Z.).

[1]M. Mandal and R. R. Breaker, Nat. Rev. Mol. Cell Biol. **5**, 451 (2004).
[2]W. C. Winkler and R. R. Breaker, ChemBioChem **4**, 1024 (2003).
[3]A. G. Vitreschak, D. A. Rodionov, A. A. Mironov, and M. S. Gelfand, Trends Genet. **20**, 44 (2004).

[4]A. Serganov and D. J. Patel, Curr. Opin. Struct. Biol. **22**, 279 (2012).
[5]A. Serganov and E. Nudler, Cell **152**, 17 (2013).
[6]M. Mandal, B. Boese, J. E. Barrick, W. C. Winkler, and R. R. Breaker, Cell **113**, 577 (2003).
[7]A. Villa, J. Wöhnert, and G. Stock, Nucleic Acids Res. **37**, 4774 (2009).
[8]L. Huang, A. Serganov, and D. J. Patel, Mol. Cell **40**, 774 (2010).
[9]M. Mandal, M. Lee, J. E. Barrick, Z. Weinberg, G. M. Emilsson, W. L. Ruzzo, and R. R. Breaker, Science **306**, 275 (2004).
[10]W. C. Winkler, A. Nahvi, N. Sudarsan, J. E. Barrick, and R. R. Breaker, Nat. Struct. Biol. **10**, 701 (2003).
[11]W. Huang, J. Kim, S. Jha, and F. Aboul-ela, Nucleic Acids Res. **37**, 6528 (2009).
[12]S. Li and R. R. Breaker, Nucleic Acids Res. **41**, 3022 (2013).
[13]J. A. Liberman and J. E. Wedekind, Wiley Interdiscip. Rev.: RNA **3**, 369 (2012).
[14]W. C. Winkler, S. Cohen-Chalamish, and R. R. Breaker, Proc. Natl. Acad. Sci. U. S. A. **99**, 15908 (2002).
[15]R. Rieder, K. Lang, D. Graber, and R. Micura, ChemBioChem **8**, 896 (2007).
[16]S. D. Gilbert, R. K. Montange, C. D. Stoddard, and R. T. Batey, Cold Spring Harbor Symp. Quant. Biol. **71**, 259 (2006).
[17]F. J. Grundy and T. M. Henkin, Mol. Microbiol. **30**, 737 (1998).
[18]R. T. Batey, S. D. Gilbert, and R. K. Montange, Nature **432**, 411 (2004).
[19]J.-C. Lin and D. Thirumalai, J. Am. Chem. Soc. **135**, 16641 (2013).
[20]A. Reining, S. Nozinovic, K. Schlepckow, F. Buhr, B. Fürtig, and H. Schwalbe, Nature **499**, 355 (2013).
[21]M. T. Cheah, A. Wachter, N. Sudarsan, and R. R. Breaker, Nature **447**, 497 (2007).
[22]K. Neupane, H. Yu, D. A. N. Foster, F. Wang, and M. T. Woodside, Nucleic Acids Res. **39**, 7677 (2011).
[23]U. D. Priyakumar and A. D. MacKerell, Jr., J. Mol. Biol. **396**, 1422 (2010).
[24]Z. Gong, Y. Zhao, C. Chen, and Y. Xiao, J. Biomol. Struct. Dyn. **29**, 403 (2011).
[25]J.-F. Lemay and D. A. Lafontaine, RNA **13**, 339 (2007).
[26]J.-F. Lemay, G. Desnoyers, S. Blouin, B. Heppell, L. Bastet, P. St-Pierre, E. Massé, and D. A. Lafontaine, PLoS Genet. **7**, e1001278 (2011).
[27]J. Noeske, C. Richter, M. A. Grundl, H. R. Nasiri, H. Schwalbe, and J. Wöhnert, Proc. Natl. Acad. Sci. U. S. A. **102**, 1372 (2005).
[28]M. Mandal and R. R. Breaker, Nat. Struct. Mol. Biol. **11**, 29 (2004).
[29]S. D. Gilbert, C. D. Stoddard, S. J. Wise, and R. T. Batey, J. Mol. Biol. **359**, 754 (2006).
[30]J.-C. Lin and D. Thirumalai, J. Am. Chem. Soc. **130**, 14080 (2008).
[31]J.-C. Lin, C. Hyeon, and D. Thirumalai, Phys. Chem. Chem. Phys. **16**, 6376 (2014).
[32]J. K. Wickiser, M. T. Cheah, R. R. Breaker, and D. M. Crothers, Biochemistry **44**, 13404 (2005).
[33]C. D. Stoddard, S. D. Gilbert, and R. T. Batey, RNA **14**, 675 (2008).
[34]O. Allnér, L. Nilsson, and A. Villa, RNA **19**, 916 (2013).
[35]J.-F. Lemay, J. C. Penedo, R. Tremblay, D. M. J. Lilley, and D. A. Lafontaine, Chem. Biol. **13**, 857 (2006).
[36]J. Noeske, J. Buck, B. Fürtig, H. R. Nasiri, H. Schwalbe, and J. Wöhnert, Nucleic Acids Res. **35**, 572 (2007).
[37]O. M. Ottink, S. M. Rampersad, M. Tessari, G. J. R. Zaman, H. A. Heus, and S. S. Wijmenga, RNA **13**, 2202 (2007).
[38]G. Quarta, K. Sin, and T. Schlick, PLoS Comput. Biol. **8**, e1002368 (2012).
[39]T. N. Wong, T. R. Sosnick, and T. Pan, Proc. Natl. Acad. Sci. U. S. A. **104**, 17995 (2007).
[40]F. R. Kramer and D. R. Mills, Nucleic Acids Res. **9**, 5109 (1981).
[41]P. Zhao, W. Zhang, and S.-J. Chen, J. Chem. Phys. **135**, 245101 (2011).
[42]J. K. Wickiser, W. C. Winkler, R. R. Breaker, and D. M. Crothers, Mol. Cell **18**, 49 (2005).
[43]K. L. Frieda and S. M. Block, Science **338**, 397 (2012).
[44]P. Zhao, W. B. Zhang, and S. J. Chen, Biophys. J. **98**, 1617 (2010).
[45]W. Zhang and S.-J. Chen, Biophys. J. **90**, 765 (2006).
[46]J. Chen, S. Gong, Y. Wang, and W. Zhang, J. Chem. Phys. **140**, 025102 (2014).
[47]J. Chen and W. Zhang, J. Chem. Phys. **137**, 225102 (2012).
[48]W. J. Greenleaf, K. L. Frieda, D. A. N. Foster, M. T. Woodside, and S. M. Block, Science **319**, 630 (2008).
[49]N. Komissarova, J. Becker, S. Solter, M. Kireeva, and M. Kashlev, Mol. Cell **10**, 1151 (2002).
[50]I. Gusarov and E. Nudler, Mol. Cell **3**, 495 (1999).
[51]L. Lubkowska, A. S. Maharjan, and N. Komissarova, J. Biol. Chem. **286**, 31576 (2011).
[52]T. Xia, J. Santalucia, Jr., M. E. Burkard, R. Kierzek, S. J. Schroeder, X. Jiao, C. Cox, and D. H. Turner, Biochemistry **37**, 14719 (1998).
[53]D. H. Mathews, J. Sabina, M. Zuker, and D. H. Turner, J. Mol. Biol. **288**, 911 (1999).
[54]I. Artsimovitch, V. Svetlov, L. Anthony, R. R. Burgess, and R. Landick, J. Bacteriol. **182**, 6027 (2000).
[55]R. A. Mooney, I. Artsimovitch, and R. Landick, J. Bacteriol. **180**, 3265 (1998).
[56]L. E. Johansen, P. Nygaard, C. Lassen, Y. AgersΦ, and H. H. Saxild, J. Bacteriol. **185**, 5200 (2003).
[57]U. Vogel and K. F. Jensen, J. Biol. Chem. **272**, 12265 (1997).
[58]S. F. Tolić-Nørrelykke, A. M. Engh, R. Landick, and J. Gelles, J. Biol. Chem. **279**, 3292 (2004).
[59]A. Haller, M. F. Soulière, and R. Micura, Acc. Chem. Res. **44**, 1339 (2011).
[60]I. Artsimovitch and R. Landick, Proc. Natl. Acad. Sci. U. S. A. **97**, 7090 (2000).
[61]N. Komissarova and M. Kashlev, Proc. Natl. Acad. Sci. U. S. A. **95**, 14699 (1998).
[62]J. A. Monforte, J. D. Kahn, and J. E. Hearst, Biochemistry **29**, 7882 (1990).
[63]D. E. Draper, RNA **10**, 335 (2004).
[64]D. E. Draper, D. Grilley, and A. M. Soto, Annu. Rev. Biophys. Biomol. Struct. **34**, 221 (2005).
[65]D. Grilley, A. M. Soto, and D. E. Draper, Proc. Natl. Acad. Sci. U. S. A. **103**, 14003 (2006).
[66]Z. J. Tan and S. J. Chen, Biophys. J. **99**, 1565 (2010).
[67]Z. J. Tan and S. J. Chen, Biophys. J. **101**, 176 (2011).