

Sequence-dependent Kinetic Model for Transcription Elongation by RNA Polymerase

Lu Bai, Alla Shundrovsky and Michelle D. Wang*

Department of Physics
Laboratory of Atomic and Solid
State Physics, Cornell
University, Ithaca, NY 14853
USA

We present a kinetic model for the sequence-dependent motion of RNA polymerase (RNAP) during transcription elongation. For each NTP incorporation, RNAP has a net forward translocation of one base-pair along the DNA template. However, this process may involve the exploration of back-tracked and forward-tracked translocation modes. In our model, the kinetic rates for the reaction pathway, calculated based on the stabilities of the transcription elongation complex (TEC), necessarily lead to sequence-dependent NTP incorporation rates. Simulated RNAP elongation kinetics is in good agreement with data from transcription gels and single-molecule studies. The model provides a kinetic explanation for well-known back-tracked pauses at transcript positions with unstable TECs. It also predicts a new type of pause caused by an energetically unfavorable transition from pre to post-translocation modes.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: RNA polymerase; transcription; kinetics; sequence-dependence; model

*Corresponding author

Introduction

Transcription is the first step in gene expression and a major point of gene regulation. Transcription elongation does not proceed at a uniform rate. In particular, RNA polymerase (RNAP) tends to dwell transiently at certain template positions known as pause sites. In recent years, much progress has been made in elucidating the mechanism of transcription elongation (for reviews, see von Hippel¹ and von Hippel & Pasmann²). Biochemical assays have led to the suggestion that transcription pausing results from misalignment of the RNA 3' end with the RNAP active site.^{3–5} Theoretical work has provided additional insights. Yager & von Hippel⁶ pioneered a static sequence-dependent thermodynamic analysis of transcription elongation complex (TEC) stability of *Escherichia coli* RNAP, and suggested that a more complete description of transcription would require kinetic analysis. Matsuzaki *et al.*⁷ showed that a simple kinetic model with a single rate-limiting step for each NTP incorporation cycle could not adequately describe the elongation rates of RNA polymerase III. Subsequently a number of models

have been proposed to describe each NTP incorporation cycle as multiple-step or even multiple-branch reactions.^{8–10} One notable work among these is a kinetic model proposed by Guajardo & Sousa⁸ that clearly points out the importance of RNAP translational motion in addition to the chemical catalysis process in transcription kinetics. Although these models have already shed much light on the general transcription mechanism, there is still no adequate explanation as to why misalignments of the RNA 3' end are correlated with specific sequences so as to induce pausing, and a theoretical basis for the sequence-dependent kinetics of transcription remains obscure.

The sequence-dependent kinetic model we present here combines the essence of the works of Yager & von Hippel⁶ and Guajardo & Sousa⁸ and extends them by constructing a quantitative sequence-dependent transcription energy landscape and performing a full kinetic analysis. This model not only provides detailed explanations of many existing experimental facts, but also predicts kinetic properties of RNAP on various DNA sequences. We expect this model to have broad predictive power, and to be readily extendable beyond transcription elongation.

RNAP Kinetic Model

Our RNAP kinetic model is based on the general

Abbreviations used: RNAP, RNA polymerase; TEC, transcription elongation complex; ss, single-stranded; ds, double-stranded.

E-mail address of the corresponding author:
mdw17@cornell.edu

thermal ratchet model for processive molecular motors: its translocations are thermally activated, and the net motion is forward biased by the RNA polymerization reaction.

Transcription reaction pathway

Figure 1(a) is a schematic representation of the transcription reaction pathway on which our model is based. RNAP, DNA, and the nascent RNA transcript form a TEC, in which a transcription bubble has been opened inside the RNAP with the

nascent RNA forming an RNA–DNA hybrid with the single-stranded (ss) template DNA. During elongation, RNAP can slide back and forth on the DNA template, forming various configurations,^{3,8} which we refer to as translocation modes. Our modes 0 and 1 are conventionally referred to in the literature as the pre and post-translocation states, respectively. Only when the TEC is in translocation mode 1 is the RNAP’s active site adjacent to the 3’ end of the RNA, so that the next nucleotide can bind and be incorporated. A nucleotide incorporation brings the TEC to translocation mode 0 and the TEC

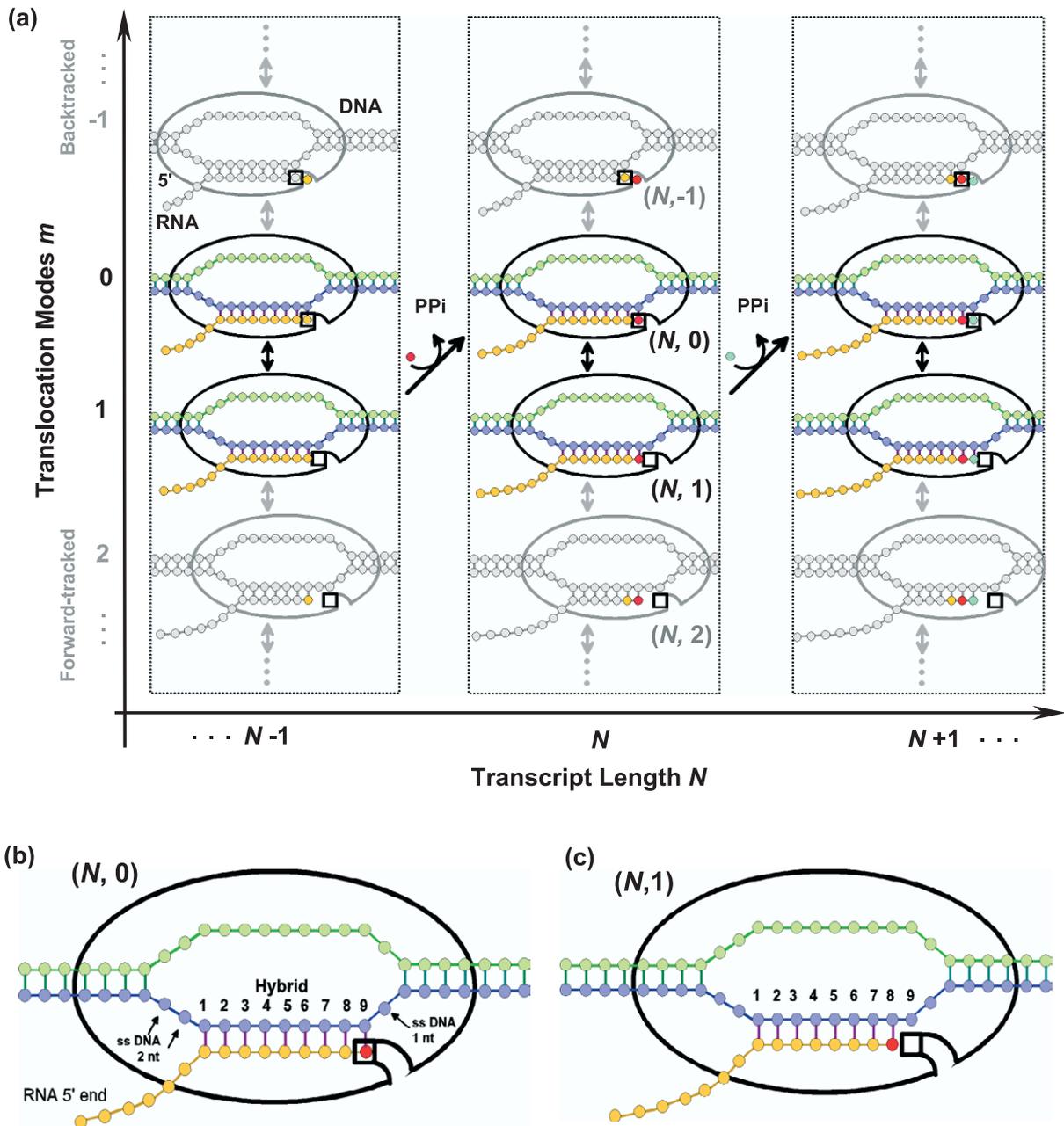


Figure 1. Schematic representations of the reaction pathways. (a) Transcription elongation pathways. Each oval-shaped cartoon represents a TEC configuration and its black box represents the active site of RNAP. The DNA template is stationary throughout the Figure; the RNAP is moving. $(N,0)$ and $(N,1)$ states are colored to indicate that they are indispensable for elongation and are part of the main pathway. (b) and (c) The structures of TECs in the $(N,0)$ and $(N,1)$ states, respectively (also see Table 1).

must move forward on the DNA template by 1 bp (returning to mode 1) to allow incorporation of the next NTP. Translocation between modes 0 and 1, together with NTP binding and catalysis, are indispensable for transcription elongation and thus belong to the main reaction pathway. On the other hand, RNAP can also enter branched pathways by back-tracking (modes $-1, -2, \dots$) or forward-tracking (modes $2, 3, \dots$). In order to simulate the kinetic details of transcription, it is necessary to quantify all reaction rates for both the main and branched pathways. In our model, these rates depend strongly on the stability of the TEC and therefore on the nucleic acid sequence within the TEC complex.

TEC state energy

Each combination of transcript length N and translocation mode m defines a TEC state (N, m) . Its state energy $\Delta G_{N,m}$ is a measure of the TEC stability and is represented by the standard free energy of formation of the complex from its isolated components:⁶

$$\begin{aligned} \Delta G_{N,m} = & \Delta G_{N,m; \text{DNA bubble}} \\ & + \Delta G_{N,m; \text{RNA-DNA hybrid}} \\ & + \Delta G_{N,m; \text{RNAP binding}} \end{aligned} \quad (1)$$

The first and second terms, respectively, are the free energy changes involved in the formation of the DNA bubble and the RNA–DNA hybrid. These two terms are clearly sequence-dependent and are calculated using a nearest-neighbor model with energy values at 24 °C^{11,12} under the assumption⁶ that base-pairing energy in a TEC is comparable to that measured in solution. For a TEC in mode 1 and forward-tracked modes, the unpaired single-stranded template DNA nucleotide immediately adjacent to the 3' end of the RNA is likely to orient itself with respect to the RNA–DNA hybrid inside the RNAP main channel and thereby contribute to the overall stability of the TEC. In the absence of

measured values for this additional energy, we approximate it as half of the free energy change that would occur if this DNA nucleotide were to pair with its corresponding RNA nucleotide (see Materials and Methods for a detailed explanation of this assumption). The third term in the state energy expression corresponds to interactions between the RNAP and the nucleic acids, and is treated as a sequence-independent constant (taken to be 0 in our case) similar to the treatment by Yager & von Hippel.⁶ Our model currently does not consider the energy contributed by possible formation of RNA secondary structures.

According to its definition, $\Delta G_{N,m}$ depends on the template DNA sequence as well as the structure of the TEC. The TEC structure is characterized by three parameters: the number of DNA base-pairs in the transcription bubble, the number of base-pairs in the RNA–DNA hybrid, and the number of ssDNA nucleotides between the active site and the downstream dsDNA (Figure 1(b) and (c)). We currently use the TEC structure of *E. coli* RNAP, although our formalism is not restricted to this polymerase. Table 1 (Experimental range) shows values obtained from established biochemical studies.^{13–15} For our calculations, we chose a set of values (Table 1, Values used here) that are within these ranges.

As first pointed out by Yager & von Hippel,⁶ we also found that the calculated $\Delta G_{N,m}$ values have a broad distribution over a typical long DNA sequence. For instance, the value of $\Delta G_{N,0}$ over the first 500 nucleotides on the pKA2 template (see Materials and Methods) varies between $5 k_B T$ and $22 k_B T$, where $k_B T$ is the thermal energy. As shown in the following sections, this large dynamic range has important implications and naturally leads to dramatically different kinetic behaviors of RNAP at different template positions.

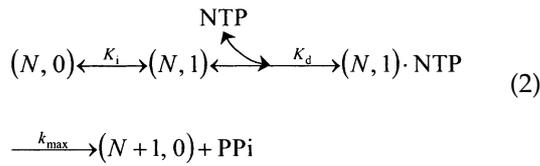
The main reaction pathway

The main reaction pathway is modeled as a three-step reaction:

Table 1. Values of model parameters

	DNA bubble (bp)	RNA–DNA hybrid (bp)	Downstream ssDNA (nt)
Experimental range ^{13–15}	12–14	8–9	1–2
Value used here	12	9	1
Predicted range	9–16	8–10	0–4
Parameter description	Symbol and value		
Rate of NTP catalysis and PPi release ^a	$k_{\text{max}} = 24.7 \pm 1.9 \text{ s}^{-1}$		
Equilibrium dissociation constant for a complementary NTP ^a	$K_{d,c} = 15.6 \pm 3.9 \text{ } \mu\text{M}$		
Equilibrium dissociation constant for a non-complementary NTP	$K_{d,nc} = 2 \times 10^4 \text{ } \mu\text{M}$		
Rate constant pre-factor	$k_0 = 1.0 \times 10^9 \text{ s}^{-1}$		
Energy barrier height of back-tracked modes ^a	$\Delta G_b^\ddagger = 46.2 \pm 0.6 k_B T$		
Energy barrier height between modes 1 and 2	$\Delta G_{1 \leftrightarrow 2}^\ddagger = 40.0 k_B T$		
Forward-tracking energy barrier slope	$s_f = 3.1 k_B T/\text{nt}$		

^a Tuned parameter.



Here, translocations between modes 0 and 1, as well as NTP binding, are assumed to follow rapid equilibrium kinetics. The translocation equilibrium assumption was suggested and used by Guajardo & Sousa.⁸ The equilibrium assumption for NTP binding has also been used extensively in RNAP elongation kinetic studies.^{8,16,17} Similar assumptions for DNA polymerases are supported by biochemical studies.^{18–20} The equilibrium assumption for NTP binding is more likely to be accurate at high NTP concentrations. If NTP binding is a diffusion-limited process and the diffusion coefficient of NTP in the RNAP secondary channel is comparable to that in bulk solution, the equilibrium could be well established at sub-nanomolar concentrations of NTP. Thus, it is reasonable to assume that equilibrium is established under our experimental conditions of $\geq 50 \mu\text{M}$ NTPs.

The main reaction pathway above is equivalent to Michaelis–Menton enzyme kinetics in the presence of a competitive inhibitor. The equilibrium constant for translocations between $(N,0)$ and $(N,1)$ is solely determined by their state energies:

$$K_i = \exp[(\Delta G_{N,1} - \Delta G_{N,0})/k_B T]$$

leading to the apparent equilibrium dissociation constant for NTP binding:

$$K'_d = K_d \{1 + \exp[(\Delta G_{N,1} - \Delta G_{N,0})/k_B T]\}$$

The irreversible step with the apparent rate k_{\max} value includes NTP catalysis, PPi release, and possibly other steps that occur after NTP binding and before the subsequent RNAP translocation. Among these steps, the NTP catalysis step is likely to be too fast to limit k_{\max} based on kinetic measurements of both RNA and DNA polymerases.^{17,18,21} Currently, we neglect pyrophosphorolysis due to the slow rate of the reverse catalysis under our typical transcription conditions of very low PPi concentration.

The overall reaction rate of the main pathway can be expressed as

$$k_{\text{main}} = \frac{k_{\max}[\text{NTP}]}{K_d \{1 + \exp[(\Delta G_{N,1} - \Delta G_{N,0})/k_B T]\} + [\text{NTP}]} \quad (3)$$

where k_{\max} and K_d are fitting parameters in the model and are tuned according to transcription velocity data from single-molecule studies (see Materials and Methods). Note that K_d^{-1} reflects the affinity of an incoming NTP for the active site. If an NTP is complementary to the template DNA then $K_d = K_{d,c}$; otherwise $K_d = K_{d,nc}$, where $K_{d,nc}$ should be much greater than $K_{d,c}$ to account for the high fidelity of transcription (see Materials and

Methods). In the current model, we also assume that the values of $K_{d,c}$, $K_{d,nc}$, and k_{\max} are independent of NTP type.

The branched reaction pathways

Translocations in the branched reaction pathways cannot be assumed to be in equilibrium. An assumption of equilibrium would give predictions of extensive back-tracking leading to frequent arrest and extensive forward-tracking leading to frequent premature termination during continuous elongation, neither of which is observed experimentally.^{22,23} The idea that translocations to and among back-tracked states are, in general, not in equilibrium is also supported by experimental evidence that back-tracking occurs much more slowly than active elongation.^{3,4} A possible reason that back-tracked energy barriers would be high is steric hindrance for the 3' end of the RNA to thread into the RNAP's secondary channel. Therefore these translocation rates, rather than equilibrium constants, must be determined for the branched reaction pathways. According to Arrhenius kinetics, the rate of translocation from a given mode (N,m) to an adjacent mode $(N,m \pm 1)$, $k_{N,m \rightarrow m \pm 1}$, depends on the difference between the height of the activation barrier ($\Delta G_{N,m \leftrightarrow m \pm 1}^+$) and the energy of the initial mode ($\Delta G_{N,m}$):

$$k_{N,m \rightarrow m \pm 1} = k_0 \exp[-(\Delta G_{N,m \leftrightarrow m \pm 1}^+ - \Delta G_{N,m})/k_B T] \quad (4)$$

where k_0 is a pre-factor constant. $\Delta G_{N,m}$ and $\Delta G_{N,m \leftrightarrow m \pm 1}^+$ can be represented in a schematic energy landscape of translocation for a given transcript size N , as shown in Figure 2(a). The energy minima represent the free energies of various translocation modes $\Delta G_{N,m}$, which are separated by energy barriers $\Delta G_{N,m \leftrightarrow m \pm 1}^+$. The barrier between modes 0 and 1 in the energy landscape is represented by a dotted curve, since it is unnecessary to know its value as long as this barrier is low enough to ensure equilibrium between these two modes. However, back-tracking and forward-tracking barriers need to be quantified. Due to the lack of direct measurements, these barrier heights were tuned from transcription gel data after some simplifying assumptions (see Materials and Methods).

Results and Discussion

Table 1 summarizes all the parameters used in this model and lists a set of parameter values valid for our experimental conditions. Note that only three parameters were tuned (Figure 2(b)–(d); also see Materials and Methods). The remaining parameters were either dependent parameters, or were set based on experimental considerations (see Materials and Methods). It is important to note that the three tuned parameters constitute a very small set given the great complexity of the transcription

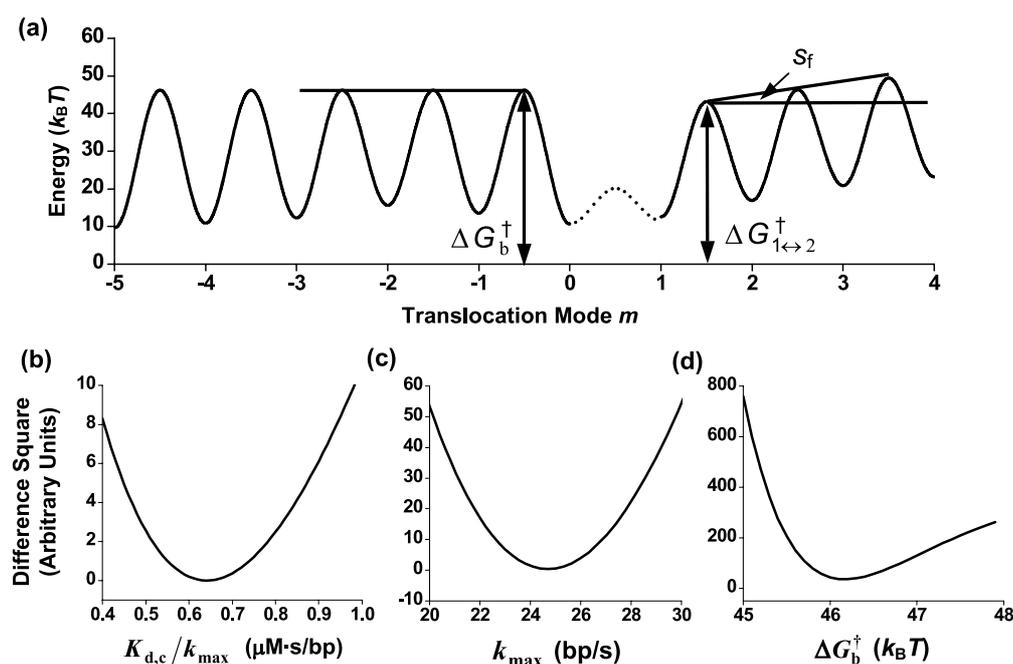


Figure 2. Schematic representation of the translocation energy landscape and tuned model parameters. (a) A typical translocation energy landscape at a given transcript length N . The parameters used to construct the landscape (see Table 1) are indicated on the Figure, in which ΔG_b^\dagger is a tuned parameter in the model. (b)–(d) Plots of the square of the difference between the simulated and experimental data as a function of $K_{d,c}/k_{max}$ (b), k_{max} (c) and ΔG_b^\dagger (d) (see Materials and Methods for details). These plots only focus on the part of parameter space where each difference is close to its minimum. Much wider ranges of parameters were explored to ensure the uniqueness of these minima.

kinetics arising from different DNA sequences. The uncertainties of tuned parameters shown in Table 1 also provide estimates of the valid ranges of these parameters, i.e. outside these parameter ranges, simulations start to deviate significantly from experimental results. Once these parameters were tuned, the model was established. The model could then be used to make a number of successful and insightful predictions.

Comparison of the model with experimental data

We simulated transcription gels on various DNA sequences and some examples are shown in Figures 3–5. The good agreement between the simulated and experimental gels indicates that the model has captured much of the essence of the sequence-dependent transcription elongation kinetics.

Figure 3(a) shows experimental and simulated transcription gels on the pTS146 template, which contains a well-characterized $\Delta tR2$ sequence with two distinctive back-tracked pause sites^{3,4,24,25} (Figure 3(b)). The only information from the $\Delta tR2$ sequence that was used to tune the model was the overall probability of pausing at the two pause sites (see Materials and Methods). This information was used to tune the back-tracked barrier height ΔG_b^\dagger , which, in the model, is the same for every template position and therefore affects their back-tracking probabilities in the same way. It is important to note that the locations of pauses, the relative probability of pausing between the two pause sites within the

$\Delta tR2$ sequence, and in fact the relative probability of pausing between any two sites on the template were not tunable in the simulated gel and were solely determined by the sequence-dependent state energies (see Materials and Methods). The model naturally predicts positions +83 and +84 to be the strongest back-tracked pause sites on the template (Figure 3(a)). In addition, when the concentration of UTP was increased from 60 μM to 210 μM , the model predicted the correct ratio of pause probabilities at each of the two sites (Figure 3(e)).

Figure 4 is a comparison of experimental and simulated gels on a DNA template that was not used to tune any model parameters. The template contains a well-characterized $\lambda tR1$ sequence,^{26,27} which is known to have three pause clusters labeled 1, 2, and 3, with pausing at cluster 2 likely being hairpin-dependent.²⁸ Since our model currently does not consider RNA hairpins, we expect that the model will not be able to make an accurate prediction for cluster 2, but should be able to predict pausing at clusters 1 and 3. Indeed the simulated gel correctly predicted pause locations within the experimental uncertainties in clusters 1 and 3. Both experimental and simulated gels also show that the pause durations decrease with an increase in NTP concentration, although some discrepancy remains in the actual durations of the pauses (see discussion below). Our model identified pauses within clusters 1 and 3 as primarily belonging to a new type of pause as discussed below in detail.

Figure 5 is another example of a comparison of experimental and simulated gels on another DNA

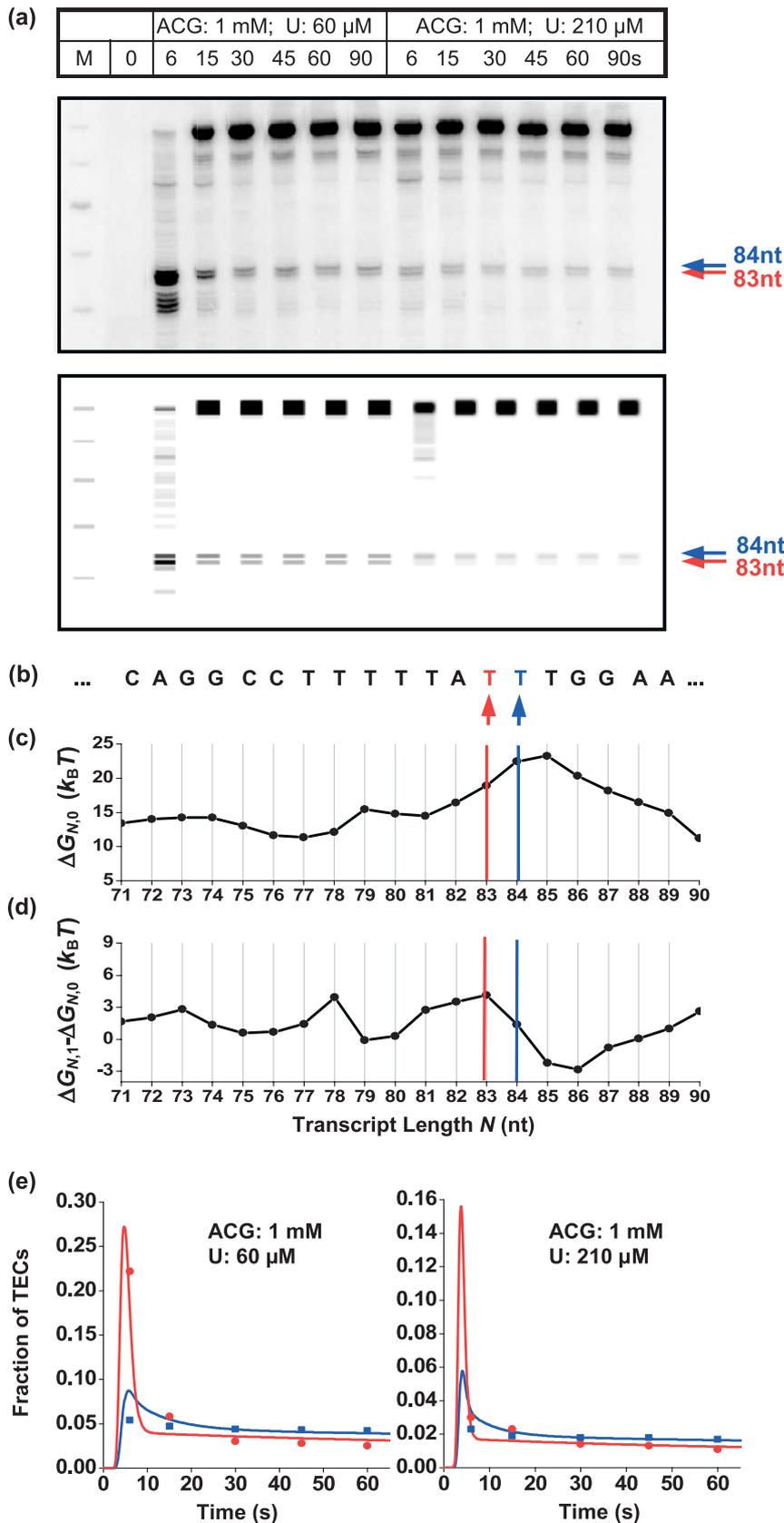


Figure 3. Simulation of the transcription gel on the pTS146 template. (a) Transcription gel on the pTS146 template containing the Δ tR2 pause sequence and its simulation. Reactions were carried out at two different UTP concentrations as indicated. The locations of the two most prominent pauses at positions +83 (red) and +84 (blue) on the experimental gel are also correctly predicted in the simulated gel. The same color system applies to (b)–(e). (b) The Δ tR2 pause sequence with the two pause sites indicated. (c) and (d) $\Delta G_{N,0}$ and $(\Delta G_{N,1} - \Delta G_{N,0})$ versus transcript length N . TECs at positions +83, +84, and +85 have a relatively high probability of back-tracking due to their high $\Delta G_{N,0}$ values. However, for the +85 site, dwell time in the back-tracked states is too short to be detectable. (e) The fraction of TECs at positions +83 and +84 as a function of time at two different UTP concentrations. The red (+83) and blue (+84) dots are experimental data obtained by line scans of the transcription experimental gel in (a), and the corresponding simulations are shown as red and blue curves.

template (pKA2) that was not used to tune any model parameters. Within the uncertainties of the experimental data, there is very good agreement between the experimental and simulated gels in

terms of the overall rate of RNA synthesis and locations of major pause sites. The model also indicates that the mechanism of pausing at those pause sites is due to RNAP back-tracking. Given

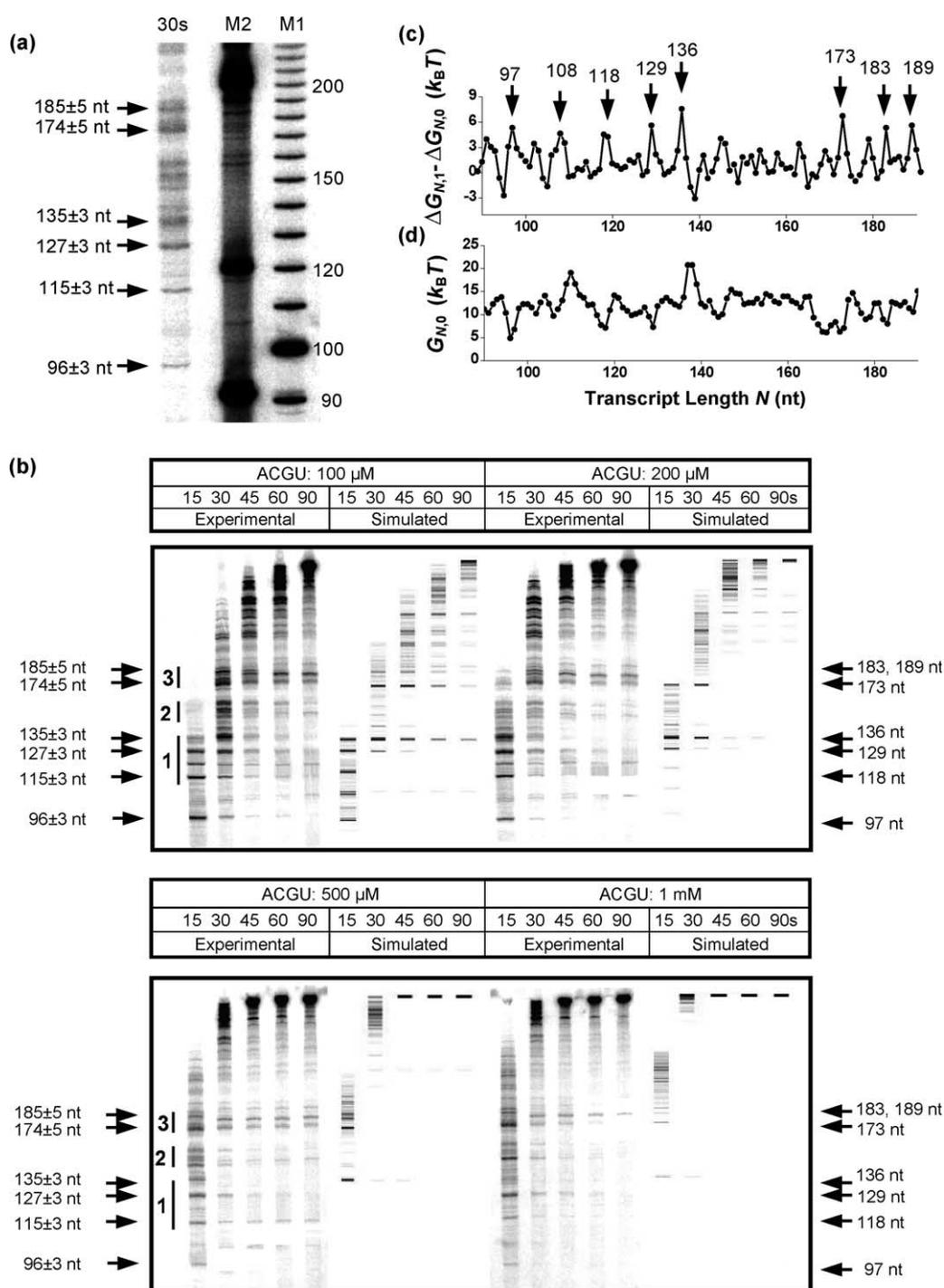


Figure 4. Simulation of the transcription gel on the λtR1 template. (a) Prominent pause positions within the λtR1 sequence on a biochemical gel. The left lane shows the transcript size distribution after 20 s of transcription on this template at 200 μM NTP. The most prominent pauses on this gel are indicated by arrows on the left and their positions were determined by comparison with both a 10 bp DNA ladder (M1) and an RNA ladder (M2). M2 was generated by runoff transcription using different lengths of PCR products of the λtR1 template, producing 90, 120, and 200 nt transcript sizes. Comparison between M1 and M2 shows that the DNA ladder can indeed be used to estimate the RNA transcript size. The uncertainties of the estimated RNA lengths are 3–5 nt, depending on the RNA length. (b) Transcription gel on the λtR1 sequence at four different NTP concentrations and its simulation. The vertical bars indicate the pause clusters 1–3 mapped from a previous biochemistry study.²⁶ The pause positions in the simulated gel (indicated by arrows on the right) agree with those of the biochemical gel (arrows on the left) within experimental uncertainties. (c) and (d) ($\Delta G_{N,1} - \Delta G_{N,0}$) and $\Delta G_{N,0}$ on the λtR1 template. The major pause positions in (a) correlate with the most prominent maxima in ($\Delta G_{N,1} - \Delta G_{N,0}$), and are identified by our model as primarily pre-translocation pauses (see the text for details).

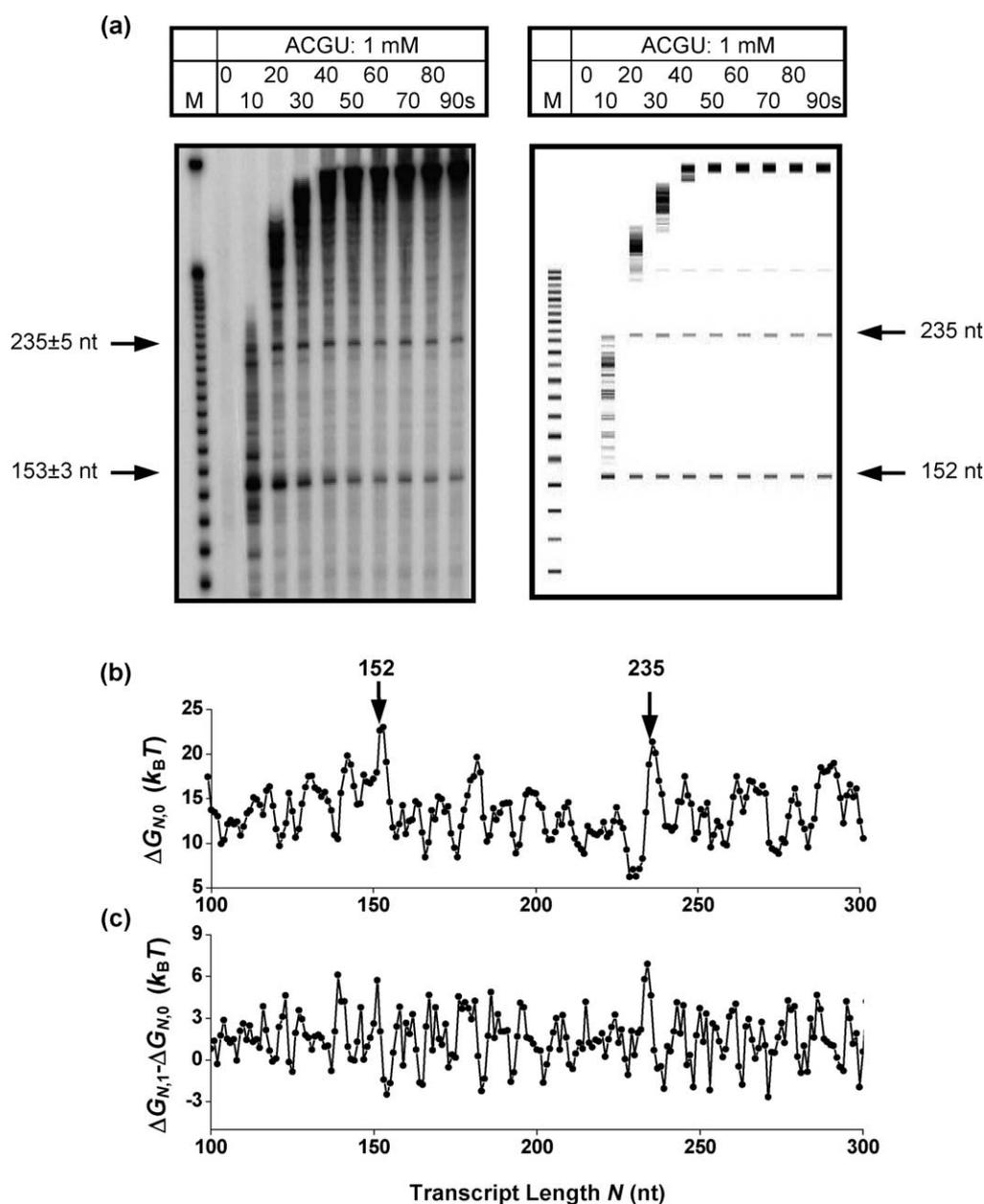


Figure 5. Simulation of the transcription gel on the pKA2 template. (a) Transcription gel on the pKA2 sequence at 1 mM NTP concentration and its simulation. The positions of the most prominent pauses identified in the biochemical gel agree with those in the simulated gel within experimental uncertainties. (b) and (c) $\Delta G_{N,0}$ and $(\Delta G_{N,1} - \Delta G_{N,0})$ on the pKA2 template. The major pause positions in (a) correlate with the prominent maxima in $\Delta G_{N,0}$, and were identified by our model as back-tracked pauses (see the text for details).

that our model presently does not consider RNA secondary structures and therefore cannot predict hairpin-dependent pauses, this agreement also indicates that hairpin-dependent pauses are not major constituents of the observed pauses on the pKA2 template.

To evaluate the agreement between experimental and simulated gels more quantitatively, we compared their intensity profiles by cross-correlation (see Materials and Methods). Three parameters characterize the normalized cross-correlation function. (1) The height of the peak near zero displacement measures the extent of correlation. Its value can be from -1 to $+1$, with $+1$ being perfect

correlation, -1 being perfect anti-correlation, and 0 being no correlation. (2) The offset of the peak from zero displacement measures the offset between the two functions that are being cross-correlated. (3) The width of the peak reflects the widths of the dominant peaks in the two functions that are being cross-correlated. An example is shown in Figure 6 with the intensity profiles (Figure 6(a) and (b)) of the 20 s lanes of the experimental and simulated pKA2 gels (Figure 5(a)) and their normalized cross-correlation function (Figure 6(c)). In Figure 6(c), the dominant peak with height ~ 0.5 indicates a good correlation between the two gels. The 1 nt offset and the ~ 3 nt width of the peak are well within the

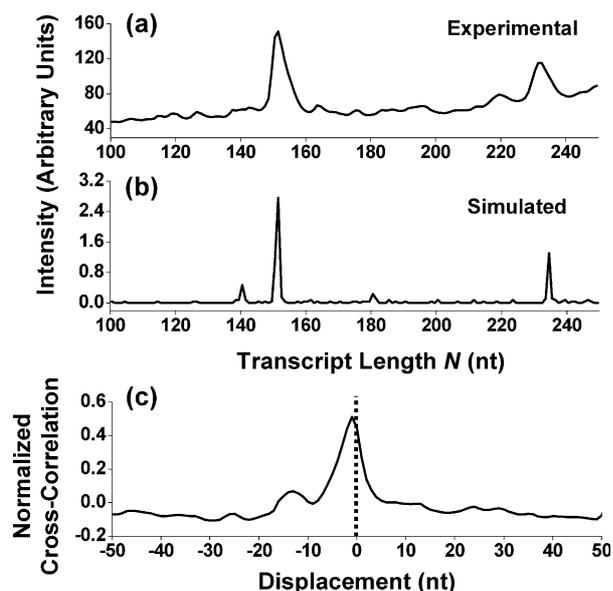


Figure 6. Normalized cross-correlation of transcription gel intensity profiles. (a) and (b) The intensity *versus* transcript length of the 20 s lane on the experimental (a) and simulated (b) pKA2 gel. (c) The normalized cross-correlation function between (a) and (b).

experimental uncertainty for position determination. Averaged over all lanes on a given gel, the correlation peak height is 0.8 for the pTS146 gel, 0.4 for the λ tR1 gel and 0.5 for the pKA2 gel. For all the gels, the offsets of the peaks are within 2 nt and the widths are about 3 nt. These data show quantitatively that the simulated gel patterns, especially the predicted major pausing sites, agree with experimental data to within experimental uncertainties.

Simulated individual RNAP motions also

correspond well with single-molecule data (Figure 7(a)). These data are from a template similar to that used for the transcription gel in Figure 5, but were necessarily taken from a different sequence section that was further from the promoter due to the time lag between the introduction of the NTPs into the sample chamber and the start of measurement. All the single molecule data used in this study were taken under low assisting force (4 pN), which would have little effect on elongation kinetics.^{29,30} Because the absolute position accuracy of the single molecule traces is ~ 100 bp,³¹ the locations of the predicted pauses cannot be verified directly. Nonetheless, other features can be compared directly. In particular, the distributions of pause duration, time between pauses, and distance between pauses show excellent agreement with those predicted by the model (Figure 7(b)–(d)).

This agreement lends further support to the model. It shows the model correctly predicts the average pausing probability, at least on this particular DNA template at this NTP concentration. Furthermore, the distributions of the time and distance between pauses follow single-exponential functions, not because pausing occurs in a random and sequence-independent manner, but because the sequence-dependent pausing sites are distributed randomly on the template and pausing probability at a given pause site is low. The distribution of pause duration follows approximately a double-exponential function, and the “fast” and “slow” components turn out to correspond to two different pausing mechanisms (see below for more details).

Mechanistic insights from the model

This model helps to identify pause sites and sheds light on the mechanism of transcription

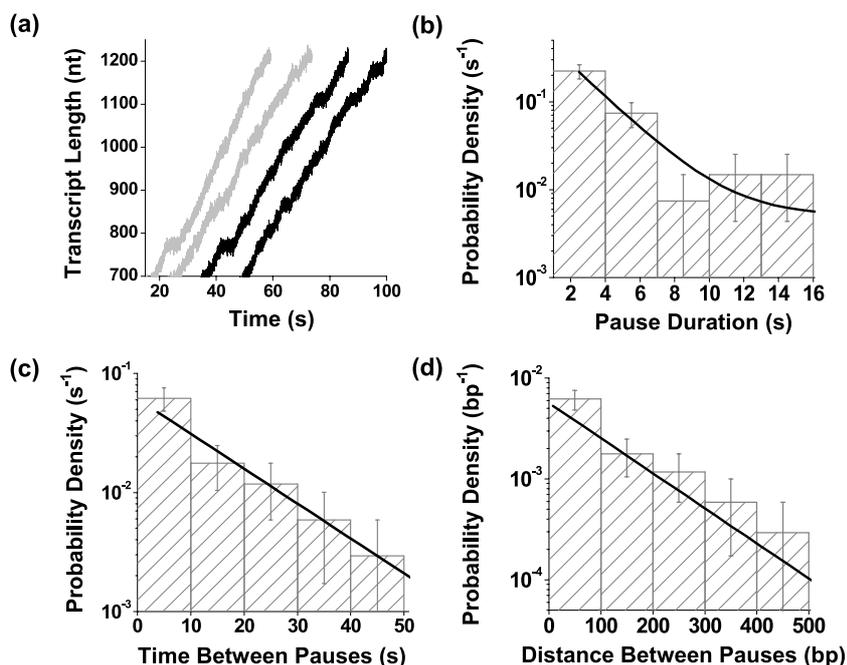


Figure 7. Comparison with single-molecule data. (a) Some examples of single-molecule traces of RNAP motion on the pRL574 template at 1 mM ATP/CTP/GTP and 50 μ M UTP under a 4 pN assisting force. Black traces are from simulations and gray traces are from experimental measurements. For clarity, the traces are offset along the time axis. (b)–(d) Comparison of experimental and predicted probability density functions of pause duration (b), time between pauses (c), and distance between pauses (d), analyzed from data such as those in (a). Experimental data are shown as vertical bars with error bars; simulations are shown as smooth curves.

pausing. Two types of pauses have been previously identified: the hairpin-induced pause⁵ and the back-tracked pause.^{3–5} As pointed out earlier, our model currently does not consider the hairpin-induced pause. However, it does consider the back-tracked pause, and also predicts a new type of pause, which we refer to as the pre-translocation pause.

The model provides a kinetic explanation for back-tracked pauses and indicates that a back-tracked pause typically occurs near an unstable TEC complex (Figures 3 and 5). Our model predicts that a back-tracked pause is induced by an unstable TEC state (Figure 8(a), green curve), rather than just

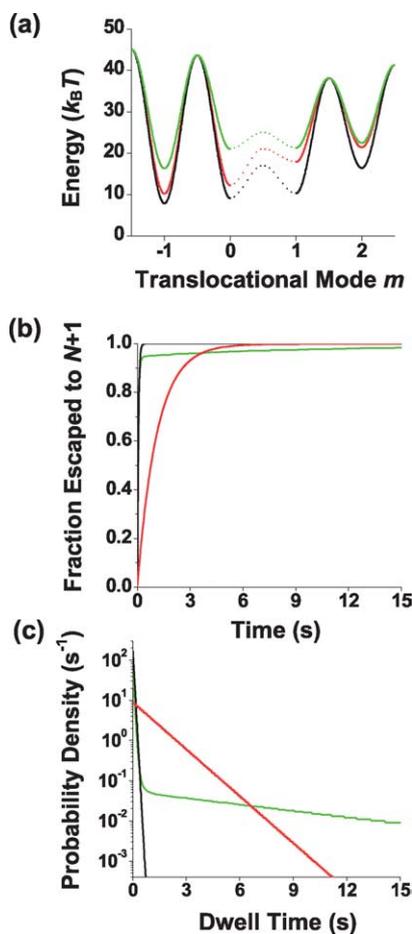


Figure 8. Mechanisms and characteristics of pauses. (a) Energy landscapes at three sites that lead to different RNAP kinetics. A back-tracked pause site (at 152 nt on pKA2; green) has a very high state energy at mode 0; while a pre-translocation pause site (at 136 nt on the λ TR1 sequence; red) has a large state energy difference between modes 1 and 0. For comparison, a non-pausing site (at 125 nt on pKA2; black) is also shown. The same color system applies to (b) and (c). (b) Time-courses for a single nucleotide incorporation at these three sites. At time 0, molecules in a population of RNAP are in state $(N,0)$. The curves show the fraction of molecules that complete a nucleotide incorporation to transcript size $N+1$ as a function of time. (c) The corresponding dwell time distributions of molecules at transcript length N before escaping to transcript size $N+1$ at these three sites.

a weak RNA–DNA hybrid (e.g. a U-stretch hybrid) although the latter is usually a special case of the former. When a TEC is in its main reaction pathway, its probability of back-tracking is determined by a competition among three different kinetic rates: back-tracking, NTP incorporation, and forward-tracking. In most cases, forward-tracking can be neglected due to its low probability of occurrence and/or short duration. To favor back-tracking over NTP incorporation, the $(N,0)$ state must be relatively unstable with a high value of $\Delta G_{N,0}$ (see equation (4)). Back-tracking is further favored over NTP incorporation if k_{main} is slow due to low NTP concentration and/or a high value of $(\Delta G_{N,1} - \Delta G_{N,0})$ (see equation (3)). However, when $\Delta G_{N,1}$ is at a local maximum, the corresponding $(\Delta G_{N,1} - \Delta G_{N,0})$ value is usually close to a local minimum. Consequently, the model predicts that the highest back-tracking probability usually occurs 1–2 bp before the template position of $\Delta G_{N,0}$ local maximum, rather than right at the local maximum (see Figure 3(c) and (d) as examples).

A back-tracked pause site that is detectable on a transcription gel requires a combination of long pause duration as well as high pause probability. For example, our model predicts three sites with high probability of back-tracking within the Δ tR2 sequence: +83, +84, and +85 nt (Figure 3). These sites are located at or immediately prior to a local maximum in $\Delta G_{N,0}$. Out of these sites, +83 and +84 pause sites were detected in both experimental and simulated gels. However, pausing at the +85 site was absent from both gels. The model indicates that at this site, its back-tracked states are so unstable that the TEC dwells only briefly in those states before returning to the main pathway.

The model predicts that at a back-tracked pause site, while a fraction of RNAP molecules is able to complete NTP incorporation without any back-tracking, some RNAP molecules back-track before the incorporation of the next NTP (Figure 8(b) and (c), green curves). Once a TEC is in a back-tracked state, it usually dwells there for a long time before returning to the $(N,0)$ state, leading to a long tail in the dwell time distribution. Our model predicts that back-tracking during continuous elongation under relatively high NTP concentration usually does not exceed 2–3 bp, and more extensive back-tracking will lead to apparent arrest over the finite experimental time-scale.

Our model also predicts a new type of pause, the pre-translocation pause. This type of pause occurs when the $(N,1)$ state is much less stable than the $(N,0)$ state (Figure 4(c), and the red curve of Figure 8(a)). Thus a pre-translocation pause occurs at a site with a large $(\Delta G_{N,1} - \Delta G_{N,0})$ and therefore a large K'_d value, resulting in a slow rate of NTP incorporation. In contrast with a back-tracked pause, the entire RNAP population is expected to follow the same, slow kinetic pathway at a pre-translocation pause site. Under high NTP concentration the pause duration at a pre-translocation pause site is typically rather short (\sim a few

seconds) (Figure 8(b) and (c), red curves), making them somewhat elusive experimentally on a transcription gel. Consequently, prominent transcription pauses identified from transcription gels are mostly back-tracked pauses, e.g. pauses at 152 nt and 235 nt on the pKA2 template (Figure 5). Nonetheless, very strong pre-translocation pauses can be observed on transcription gels, especially under low NTP concentrations, such as the well-defined pause clusters 1 and 3 on the λ tR1 sequence²⁷ (Figure 4).

Our model also provides an explanation for the short and long pause components in the pause duration histograms observed from previous single-molecule studies;^{30,31} they are attributable to pre-translocation and back-tracked pauses, respectively. The kinetics at both back-tracked and pre-translocation pause sites are sensitive to NTP concentration. For both types of pause, a lower [NTP] leads to enhanced pausing, and the dependence of pausing on [NTP] is affected strongly by the rate constants at the pause site, and is therefore sequence specific. In general, a pre-translocation pause is more sensitive to [NTP] than a back-tracked pause.

The kinetic predictions described above depend strongly on the TEC stability calculations. Since TEC stability is a strong function of the TEC structural parameter values, we investigated whether other values could also produce the same

pausing patterns. We therefore computed $\Delta G_{N,0}$ on the pKA2 template using different TEC structure parameter values, and found ranges of values that could produce the prominent pausing patterns at the correct locations within the uncertainties of the experimental values (Table 1, Predicted range; Figure 9). These predicted ranges will likely narrow further with more extensive comparison with experimental data. As shown in Table 1, our predicted ranges of the TEC structure parameter values have considerable overlap with biochemically and structurally determined values. This agreement further supports our conclusion that the majority of the observable pauses are due to unstable TECs.

One important characterization of the RNAP motor is its force-velocity relation. That is, how the transcription velocity depends on an external load exerted on the RNAP. Our model predicts a small but non-zero dependence of transcription velocity on an external force even under saturating NTP concentrations, which is inconsistent with conclusions from a number of previous single-molecule studies.^{29,30,32} However, we found transcription elongation in single molecule studies to be extremely sensitive to experimental conditions. Therefore, it is not surprising that previous single-molecule studies performed under apparently similar conditions did not produce the same transcription kinetics. A fair comparison of our

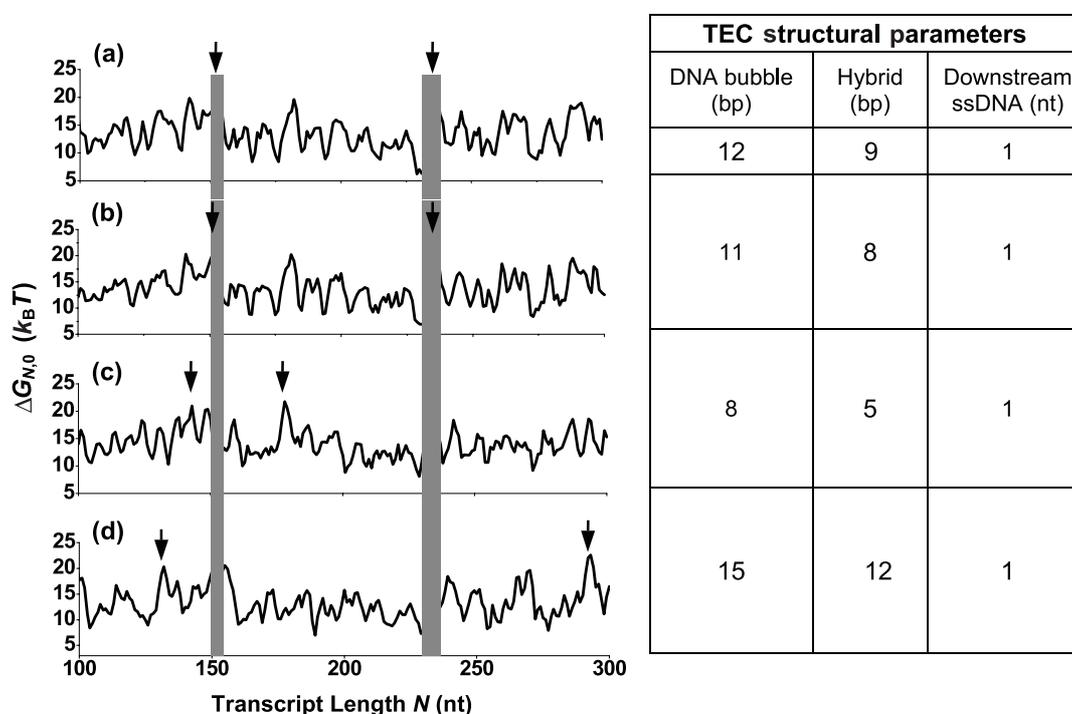


Figure 9. TEC structural parameter values determined using the pKA2 template. (a)–(d) Some examples of $\Delta G_{N,0}$ on the pKA2 template calculated with different TEC structural parameter values as indicated in the table on the right. Gray bars indicate the pause positions estimated from the experimental gel (Figure 5) with the width of the bars indicating their uncertainties. (a) When TEC structural parameter values used in the simulations presented here were used to compute $\Delta G_{N,0}$, the two highest maxima (indicated by vertical arrows) correlated well with the three expected major pause sites. (b) Good correlation still exists when $\Delta G_{N,0}$ values were calculated from TEC parameter values close to those used in (a). (c) and (d) The correlation is reduced as the parameter values deviate significantly from those used in (a).

model with experiments on the force–velocity relation requires that the experiments be performed under transcription conditions identical with those used to tune the model parameters. This will be realized in our future endeavors.

Limitations of the current model and outlook

We present a model that shows much promise in predicting transcription elongation kinetics. Nonetheless, some discrepancies remain between the simulated and experimental transcription data. Future improvements in the model will likely improve its predictive power. For example, RNA hairpins can be incorporated into this model in order to predict hairpin-dependent pauses. It is also possible that different NTPs have different K_d and k_{max} values, which were assumed to be the same for all NTPs in the current model. Incorporation of NTP-dependent parameters will likely allow more precise predictions of RNAP motions, especially at lower NTP concentrations. The current model also does not predict accurately some pause durations, possibly in part due to the assumption that the back-tracked activation barrier heights are transcript size-independent. An additional contribution to this discrepancy could arise from the uncertainties in the state energy calculations ($\sim 1 k_B T$), which are due to uncertainties in the available experimental thermodynamic values for DNA–DNA or RNA–DNA duplexes.^{11,12}

Biochemically, it has been difficult to identify unique or consensus pause sequences. Thus, an important contribution of this model is its ability to predict new pause sites, some of which may serve as regulatory signals for transcription. This model indicates that a large number of DNA sequences may produce back-tracked or pre-translocation pause sites, depending on the stability of the TEC at those sites. This work illustrates the importance of the TEC stability in transcription elongation as was first proposed by Yager & von Hippel⁶ over a decade ago, and provides a kinetic explanation of sequence-dependent RNAP motions.

Materials and Methods

Transcription assays

Bulk transcription assays and single-molecule measurements were carried out using *E. coli* RNAP at $24(\pm 1)$ °C in buffer containing 25 mM Tris–HCl (pH 8.1), 100 mM KCl, 4 mM MgCl₂, 1 mM DTT, 3% (v/v) glycerol, following protocols similar to those described.³¹ Four DNA templates were used: (1) pKA2 template composed of the λ phage 82 late promoter followed by the *rpoB* gene; (2) pRL574 template composed of the T7A1 promoter followed by the *rpoB* gene; (3) pTS146 template containing the λ phage 82 late promoter and the Δ tR2 pause sequence (tR2 terminator with truncated upstream half of the hairpin);²³ (4) λ tR1 template containing the λ phage 82 late promoter and the CA159 Δ 13 sequence²⁶ at 113–324 nt which contains the pause clusters 1–3.

Energetic contribution of the ss template DNA nucleotide immediately adjacent to the 3' end of the RNA

Previous thermodynamic studies^{33,34} of DNA or RNA duplexes showed a significant contribution of the dangling end to the overall stability of the duplex and the corresponding “dangling energy” is strongly sequence-dependent. For a DNA duplex, the average dangling energy is $0.8 k_B T$ at room temperature,³³ which is about 30% of the average energy contributed by a complete base-pair ($2.8 k_B T$).¹¹ For an RNA duplex, the average dangling energy ($0.34 k_B T$)³⁴ is about 10% of the average base-pairing energy ($3.8 k_B T$).³⁴ Therefore, it is reasonable to assume that the single-stranded template DNA nucleotide immediately adjacent to the 3' end of the RNA in the transcription bubble will contribute to the stability of the TEC, since it will likely orient itself with respect to the RNA–DNA hybrid inside the RNAP main channel. However, there are no experimental data available for the dangling energy of an RNA–DNA hybrid. Furthermore, the dangling energy in the RNAP would be expected to be different from that measured in solution because of constraints imposed by the RNAP. In the ($N,0$) or other back-tracked states, the dangling energy is likely to be small, since the single-stranded template DNA nucleotide immediately adjacent to the RNA 3' end is largely bent away from the RNA–DNA hybrid and extensively interacts with the RNAP.³⁵ In contrast, for the ($N,1$) and other forward-tracked states, the single-stranded template DNA nucleotide is brought into a large protein cavity so that its base has no direct interaction with RNAP and its orientation is fixed by constraints imposed by the phosphate backbone that links it to the adjacent upstream and downstream nucleotides in the template strand.³⁶ This oriented base thus probably has stronger interactions with the adjacent RNA–DNA hybrid than those from a dangling end in solution. Therefore, we added a dangling energy term to the ($N,1$) and other forward-tracked state energies, and assumed it to be 50% of the free energy of the corresponding terminal base-pair. This treatment is similar to that described by Yager & von Hippel,⁶ where the terminal DNA energy was taken as 50% of the complete base-pair energy.

Parameter tuning

Simulated data were generated by using Monte Carlo techniques and/or by solving a set of coupled first-order differential kinetic equations. The three tuned parameters, k_{max} , $K_{d,c}$ and ΔG_b^\ddagger , discussed in detail below, were decoupled in the tuning process and their values listed in Table 1 were obtained by the least-squares method, which minimized the difference between simulated results and the corresponding experimental data. The set of best-fit parameter values was first found by a Simplex algorithm³⁷ and later confirmed by exhaustive search over a very large range in the parameter space (over several orders of magnitude). Only one minimum was detected over this range and the uncertainty in each fitting parameter is reflected by its uncertainty, as shown in Table 1.

k_{max} , $K_{d,c}$ and $K_{d,nc}$

k_{max} and $K_{d,c}$ can be determined by the average elongation velocity when RNAP is not paused under various NTP concentrations. In principle, this velocity could be obtained from biochemical transcription gels. In

practice, transcription gels can only give rough estimates. This is because the transcript size at a given time-point on a gel shows active elongation interrupted by numerous pauses; some are long and many are somewhat short, leading to asynchronous RNAP motion. Furthermore, the probability of pausing at any given pause site is not always 100% and the escape from a pause often follows complex kinetics, leading to additional asynchrony. The resulting large spread in the RNA transcript size makes it difficult to estimate the average elongation velocity from a transcription gel. Therefore, we determined k_{\max} and $K_{d,c}$ based on single-molecule transcription measurements.

The single-molecule experimental configuration used here was the same as described,³¹ except that the microsphere was attached to the upstream DNA end in order to exert an assisting force on the RNAP. Briefly, paused transcription complexes were formed by NTP starvation on the pRL574 template using HA-tagged RNAP and immobilized on an anti-HA-coated microscope coverslip in a flow chamber. After the addition of 1 mM of all four NTPs (condition 1) or 1 mM ATP/CTP/GTP and 50 μ M UTP (condition 2), the motion of individual RNAP molecules was monitored in real time under 4 pN of assisting force using an optical trapping setup. The 4 pN force is low enough so as not to alter transcription kinetics significantly.^{29,30} The position *versus* time data on the pRL574 template from +200 nt to +1000 nt were smoothed using a 1.0 s Gaussian low-pass filter and the instantaneous transcription velocity from each trace was computed. Then, a velocity distribution was generated with equal statistical weight given to each template position. This analysis effectively removed the contribution due to pausing, since no change in template position took place during a pause. Combining velocity distributions for all traces obtained under a given NTP condition produced a well-defined non-zero velocity peak that was well fit by a single Gaussian function. The mean of this distribution is therefore a measure of the average (over the template) elongation velocity of the main pathway $\langle k_{\text{main}} \rangle_{\text{exp}}$.

Corresponding simulated single-molecule traces of transcript size *versus* time were generated using a range of parameter values for k_{\max} and $K_{d,c}$. White noise (± 10 bp standard deviation) was added to these traces to simulate thermal fluctuations under experimental conditions. These simulated traces were then similarly analyzed to obtain the average (over the template) elongation velocity from the main pathway $\langle k_{\text{main}} \rangle_{\text{sim}}$.

To tune k_{\max} and $K_{d,c}$, we conceptually take advantage of the following relation derived from equation (3):

$$\begin{aligned} & \frac{1}{k_{\text{main}}([\text{NTP}]_1)} - \frac{1}{k_{\text{main}}([\text{NTP}]_2)} \\ &= \frac{K_{d,c}}{k_{\max}} \{1 + \exp [(\Delta G_{N,1} \\ & \quad - \Delta G_{N,0})/k_B T]\} \left(\frac{1}{[\text{NTP}]_1} - \frac{1}{[\text{NTP}]_2} \right) \end{aligned} \quad (5)$$

where $k_{\text{main}}([\text{NTP}]_1)$ is the k_{main} value obtained at NTP condition 1 and $k_{\text{main}}([\text{NTP}]_2)$ is the k_{main} obtained at NTP condition 2. This relation shows that the difference between the reciprocal of the main pathway elongation rates under two different NTP conditions is solely determined by the parameter $K_{d,c}/k_{\max}$. The best-fit value of $K_{d,c}/k_{\max}$ was obtained by minimizing the difference between the simulated and experimental values of:

$$\frac{1}{\langle k_{\text{main}}([\text{NTP}]_1) \rangle} - \frac{1}{\langle k_{\text{main}}([\text{NTP}]_2) \rangle}$$

Once the best-fit value of $K_{d,c}/k_{\max}$ was obtained, we further determined k_{\max} by conceptually using the following relation:

$$\begin{aligned} \frac{1}{k_{\text{main}}([\text{NTP}])} &= \frac{1}{k_{\max}} + \frac{K_{d,c}}{k_{\max}} \{1 + \exp [(\Delta G_{N,1} \\ & \quad - \Delta G_{N,0})/k_B T]\} \frac{1}{[\text{NTP}]} \end{aligned} \quad (6)$$

The best-fit value of k_{\max} was obtained by minimizing the difference between the simulated and experimental values of $\langle k_{\text{main}} \rangle$ under both NTP conditions.

Once $K_{d,c}/k_{\max}$ and k_{\max} values were known, $K_{d,c}$ was then determined. It is worth noting that the tuned value of $K_{d,c}$ is comparable to values determined experimentally by Rhodes & Chamberlin¹⁶ although their measured values actually correspond to our $K_{d,c}^*$ value, which is greater than $K_{d,c}$. However, we simulated transcription at 24 °C whereas their experiments were carried out at 37 °C, and it is known that dissociation constants decrease with an increase in temperature. It is likely that $K_{d,c}$ is NTP type-dependent¹⁶ so the $K_{d,c}$ value determined using our method may better reflect that of UTP, since only [UTP] was varied here.

For simplicity, our model currently allows competition and incorporation of non-complementary NTPs, but it does not consider the altered incorporation kinetics after mis-incorporation.³⁸ Under this assumption, our model has little dependence on $K_{d,c}$ as long as it is much greater than $K_{d,nc}$. $K_{d,nc}$ was set to be $\sim 10^3 K_{d,c}$ so as to account for the experimentally observed $\sim 10^{-3}$ mis-incorporation rate for *in vitro* transcription.³⁸ This treatment is probably an oversimplification, since there is evidence that the difference between $K_{d,c}$ and $K_{d,nc}$ is not this large¹⁶ so that the NTP incorporation specificity does not solely result from differences in K_d , but probably also from differences in k_{\max} . This simplification would not affect the validity of the model's predictions significantly because of the low probability of mis-incorporation; even under 1 mM ATP/CTP/GTP and 50 μ M UTP concentration, less than 1% of the TECs will undergo mis-incorporation, even at a U site.

k_0 , ΔG_b^\ddagger , $\Delta G_{1 \leftrightarrow 2}^\ddagger$ and s_f

Among these four parameters, the pre-factor k_0 is not an independent parameter and was arbitrarily set to 10^9 s^{-1} . As previously mentioned, to avoid extensive back-tracking and forward-tracking in the predicted RNAP kinetics, it was necessary to assume both large back-tracked and forward-tracked barriers. To get the values for $\Delta G_{N,m \leftrightarrow m \pm 1}^\ddagger$, several assumptions were made to simplify this rather complex problem. $\Delta G_{N,m \leftrightarrow m \pm 1}^\ddagger$ was assumed to be independent of N and was represented by $\Delta G_{m \leftrightarrow m \pm 1}^\ddagger$. Because the back-tracked translocation modes all have the same average (over all transcript sizes) state energy, the back-tracked barrier heights were also assumed to be the same for all the back-tracked modes (ΔG_b^\ddagger). On the other hand, the forward-tracked translocation modes have an average state energy that increases linearly with m , with slope s_f , due to the gradual loss of the RNA-DNA hybrid. Therefore, we assumed the forward-tracked barrier heights to have the same slope of s_f (Figure 2(a)). Consequently, $\Delta G_{1 \leftrightarrow 2}^\ddagger$ was the lowest forward-tracked barrier, and s_f was computed to be 3.1 $k_B T$ per forward-tracking step.

ΔG_b^+ strongly affects the probability of RNAP pausing due to back-tracking:

$$P_{\text{backtracking}}(N) = \frac{k_{N,0 \rightarrow -1}}{k_{\text{main}}(N) + k_{N,0 \rightarrow -1}} \approx \frac{k_{N,0 \rightarrow -1}}{k_{\text{main}}(N)} \\ = \frac{k_0}{k_{\text{main}}(N)} \exp [-(\Delta G_b^+ - \Delta G_{N,0})/k_B T] \quad (7)$$

where the rate of back-tracking $k_{N,0 \rightarrow -1} \ll k_{\text{main}}(N)$ under our typical experimental conditions. However, the relative probability of back-tracking between any two sites with transcript lengths $N1$ and $N2$ on the template does not depend on ΔG_b^+ and is solely determined by the TEC free energy:

$$\frac{P_{\text{backtracking}}(N1)}{P_{\text{backtracking}}(N2)} = \frac{k_{\text{main}}(N2)}{k_{\text{main}}(N1)} \exp [(\Delta G_{N1,0} - \Delta G_{N2,0})/k_B T] \quad (8)$$

ΔG_b^+ was tuned according to the $\Delta tR2$ pause probability as a function of time under two NTP concentrations in a biochemical transcription gel (1 mM ATP/GTP/CTP and 210 μ M UTP; and 1 mM of ATP/GTP/CTP and 60 μ M UTP) (Figure 3(e)).

Forward-tracked barrier heights affect the rates and extent of forward-tracking. Some previous studies have shown that extensive forward-tracking for 5–8 nt would lead to dissociation of the TEC complex (termination).²³ In the current model, we assumed that the TEC would dissociate after RNAP forward-tracked 6 bp. It was found in the simulation that, even though the forward-tracked states were energetically unfavorable, a detectable portion of RNAP could forward-track to the dissociation point if the $\Delta G_{1 \leftrightarrow 2}^+$ value was set too low, especially for unstable TECs. This conflicts with the experimental observation that premature termination rarely happens. Once $\Delta G_{1 \leftrightarrow 2}^+$ was increased so that the predicted termination probability was negligible, the elongation kinetics were no longer sensitive to the barrier height due to the short dwell time and/or the low probability of forward-tracking. So $\Delta G_{1 \leftrightarrow 2}^+$ could not be tuned according to available RNAP elongation data, and in the current model, $\Delta G_{1 \leftrightarrow 2}^+$ was chosen to be high enough to prevent termination.

We note that although the tuned activation barriers under our assumptions work well in the simulation, this choice of activation barriers may not be unique. Our choice of activation barriers resulted in a very small number of tuning parameters and thus greatly simplified the simulation.

Uncertainties in the tuned parameters

Two factors influenced the determination of the uncertainties of the tuned parameters listed in Table 1. First, because these parameters were tuned according to certain experimental data, uncertainties in the experimental data resulted in uncertainties in the tuned parameters. Second, each parameter had an optimal range of values outside of which the simulated and experimental results started to deviate significantly. Figure 2(b)–(d) shows how the square of the difference between simulated and experimental data varied with each parameter value. The uncertainty range of a parameter was defined as the range that corresponded to twice the minimum value of the square of the difference between simulated and experimental data. For a conservative estimate of the uncertainty in a tuned

parameter, we listed in Table 1 the largest of these two sources of uncertainties for that parameter.

Normalized cross-correlation of transcription gel intensity profiles

The intensity profile was obtained by line scanning through each lane on the experimental gel using Scion Image (Scion Corporation). The horizontal axis of the profile represents the positions on the gel, which depends on the transcript length logarithmically. Using the positions of ladders, we mapped the horizontal axis to its corresponding transcript length in base-pairs (see Figure 6(a) as an example), which could directly cross-correlate with a corresponding simulated intensity profile (Figure 6(b)). We used a normalized cross-correlation function defined as:

$$C(x) = \frac{\int dx' [f(x') - \langle f \rangle][g(x + x') - \langle g \rangle]}{\sqrt{\int dx' [f(x') - \langle f \rangle]^2 \int dx' [g(x') - \langle g \rangle]^2}} \quad (9)$$

where $f(x)$ and $g(x)$ are the two functions to be cross-correlated. The result of the normalized cross-correlation is shown in Figure 6(c).

Acknowledgements

We thank Drs J. P. Sethna, J. W. Roberts, R. M. Fulbright, T. J. Santangelo, A. La Porta, K. Adelman, and S. J. Koch for stimulating scientific discussions and helpful technical advice to expedite the parameter tuning process. We thank Drs K. Adelman and T. J. Santangelo for purification of HA-tagged *E. coli* RNAP and the experimental transcription gel on the pKA2 template. We also thank Dr R. Landick for the pRL574 plasmid. M.D.W. was supported by grants from the NIH, the Beckman Young Investigator Award, the Alfred P. Sloan Research Fellow Award, and the Keck Foundation's Distinguished Young Scholar Award.

References

1. von Hippel, P. H. (1998). An integrated model of the transcription complex in elongation, termination, and editing. *Science*, **281**, 660–665.
2. von Hippel, P. H. & Pasman, Z. (2002). Reaction pathways in transcript elongation. *Biophys. Chem.* **101–102**, 401–423.
3. Komissarova, N. & Kashlev, M. J. (1997). RNA Polymerase switches between inactivated and activated states by translocating back and forth along the DNA. *Biol. Chem.* **272**, 15329–15338.
4. Komissarova, N. & Kashlev, M. J. (1997). Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3' end of the RNA intact and extruded. *Proc. Natl Acad. Sci. USA*, **94**, 1755–1760.
5. Artsimovitch, I. & Landick, R. (2000). Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl Acad. Sci. USA*, **97**, 7090–7095.
6. Yager, T. D. & von Hippel, P. H. (1991). A

- thermodynamic analysis of RNA transcript elongation and termination in *Escherichia coli*. *Biochemistry*, **30**, 1097–1118.
7. Matsuzaki, H., Kassavetis, G. A. & Geiduschek, E. P. (1994). Analysis of RNA chain elongation and termination by *Saccharomyces cerevisiae* RNA polymerase III. *J. Mol. Biol.* **235**, 1173–1192.
 8. Guajardo, R. & Sousa, R. (1997). A model for the mechanism of polymerase translocation. *J. Mol. Biol.* **265**, 8–19.
 9. Jülicher, F. & Bruinsma, R. (1998). Motion of RNA polymerase along DNA: a stochastic model. *Biophys. J.* **74**, 1169–1185.
 10. Wang, H., Eiston, T., Mogilner, A. & Oster, G. (1998). Force generation in RNA polymerase. *Biophys. J.* **74**, 1186–1202.
 11. SantaLucia, J., Allawi, H. T. & Seneviratne, P. A. (1996). Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry*, **35**, 3555–3562.
 12. Sugimoto, N., Nakano, S., Katoh, M., Matsumura, A., Nakamuta, H., Ohmichi, T. *et al.* (1995). Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry*, **34**, 11211–11216.
 13. Nudler, E., Gusarov, I., Avetissova, E., Kozlov, M. & Goldfarb, A. (1998). Spatial organization of transcription elongation complex in *Escherichia coli*. *Science*, **281**, 424–428.
 14. Korzheva, N., Mustaev, A., Kozlov, M., Malhotra, A., Nikiforov, V., Goldfarb, A. & Darst, S. A. (2000). A structural model of transcription elongation. *Science*, **289**, 619–625.
 15. Zhang, G., Campbell, E. A., Minakhin, L., Richter, C., Severinov, K. & Darst, S. A. (1999). Crystal structure of *Thermus aquaticus* core RNA polymerase at 3.3 Å resolution. *Cell*, **98**, 811–824.
 16. Rhodes, G. & Chamberlin, M. J. (1974). Ribonucleic acid chain elongation by *Escherichia coli* ribonucleic acid polymerase I. Isolation of ternary complexes and the kinetics of elongation. *J. Biol. Chem.* **249**, 6675–6683.
 17. Foster, J. E., Holmes, S. F. & Erie, D. A. (2001). Allosteric binding of nucleoside triphosphates to RNA polymerase regulates transcription elongation. *Cell*, **106**, 243–252.
 18. Johnson, K. A. (1993). Conformational coupling in DNA polymerase fidelity. *Annu. Rev. Biochem.* **62**, 685–713.
 19. Patel, S. S., Wong, I. & Johnson, K. A. (1991). Pre-steady-state kinetic analysis of processive DNA replication including complete characterization of an exonuclease-deficient mutant. *Biochemistry*, **30**, 511–525.
 20. Bryant, F. R., Johnson, K. A. & Benkovic, S. J. (1983). Elementary steps in the DNA polymerase I reaction pathway. *Biochemistry*, **22**, 3537–3546.
 21. Nedialkov, Y. A., Gong, X. Q., Hovde, S. L., Yamaguchi, Y., Handa, H., Geiger, J. H. *et al.* (2003). NTP-driven translocation by human RNA polymerase II. *J. Biol. Chem.* **278**, 18303–18312.
 22. Huang, J. & Sousa, R. (2000). T7 RNA polymerase elongation complex structure and movement. *J. Mol. Biol.* **303**, 347–358.
 23. Yarnell, W. S. & Roberts, J. W. (1999). Mechanism of intrinsic transcription termination and antitermination. *Science*, **284**, 611–615.
 24. Nudler, E., Mustaev, A., Lukhtanov, E. & Goldfarb, A. (1997). The RNA–DNA hybrid maintains the register of transcription by preventing back-tracking of RNA polymerase. *Cell*, **89**, 33–41.
 25. Gusarov, I. & Nudler, E. (1999). The mechanism of intrinsic transcription termination. *Mol. Cell.* **3**, 495–504.
 26. Hart, C. M. & Roberts, J. W. (1994). Deletion analysis of the Lambda tR1 termination region. *J. Mol. Biol.* **237**, 255–265.
 27. Lau, L. F., Roberts, J. W. & Wu, R. (1983). RNA polymerase pausing and transcript release at the λtR1 termination *in vitro*. *J. Biol. Chem.* **258**, 9391–9397.
 28. Rosenberg, M., Court, D., Shimatake, H. & Brady, C. (1978). The relationship between function and DNA sequence in an intercistronic regulatory region in phage λ. *Nature*, **272**, 414–422.
 29. Wang, M. D., Schnitzer, M. J., Yin, H., Landick, R., Gelles, J. & Block, S. M. (1998). Force and velocity measured for single molecules of RNA polymerase. *Science*, **282**, 902–907.
 30. Neuman, K. C., Abbondanzieri, E. A., Landick, R., Gelles, J. & Block, S. M. (2003). Ubiquitous transcriptional pausing is independent of RNA polymerase back-tracking. *Cell*, **115**, 437–447.
 31. Adelman, K., La Porta, A., Santangelo, T. J., Lis, J. T., Roberts, J. W. & Wang, M. D. (2002). Single molecule analysis of RNA polymerase elongation reveals uniform kinetic behaviour. *Proc. Natl Acad. Sci. USA*, **99**, 13538–13543.
 32. Forde, N. R., Izhaky, D., Woodcock, G. R., Wuite, G. L. & Bustamante, C. (2002). Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by *Escherichia coli* RNA polymerase. *Proc. Natl Acad. Sci. USA*, **99**, 11682–11687.
 33. Bommarito, S., Peyret, N. & SantaLucia, J. (2000). Thermodynamic parameters for DNA sequences with dangling ends. *Nucl. Acids Res.* **28**, 1929–1934.
 34. Serra, M. J. & Turner, D. H. (1995). Predicting thermodynamic properties of RNA. *Methods Enzymol.* **259**, 242–260.
 35. Yin, Y. W. & Steitz, T. A. (2004). The structural mechanism of translocation and helicase activity in T7 RNA polymerase. *Cell*, **116**, 393–404.
 36. Temiakov, D., Patlan, V., Anaikin, M., McAllister, W. T., Yokoyama, S. & Vassilyev, D. G. (2004). Structural basis for substrate selection by T7 RNA polymerase. *Cell*, **116**, 381–391.
 37. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992). *Numerical Recipes in C*, p. 408, Cambridge University Press, Cambridge.
 38. Erie, D. A., Hajiseyedjavadi, O., Young, M. C. & von Hippel, P. H. (1993). Multiple RNA polymerase conformations and GreA: control of the fidelity of transcription. *Science*, **262**, 867–873.

Edited by R. Ebricht

(Received 24 February 2004; received in revised form 6 August 2004; accepted 6 August 2004)