CHAPTER 9

Kinetic Modeling of Transcription Elongation

LU BAI,[a, b] ALLA SHUNDROVSKY[a, c] AND MICHELLE D. WANG[a]

[a] Department of Physics, Laboratory of Atomic and Solid State Physics, Cornell University, Ithaca, NY 14853, USA; [b] Current address: The Rockefeller University, New York, NY 10065, USA; [c] Current address: Department of Mechanical Engineering, Yale University, New Haven, CT 06511, USA

9.1 Introduction

Transcription elongation is the process by which RNA polymerase (RNAP) moves along template DNA and synthesizes a complementary RNA. During elongation, RNAP carries out a highly processive and directional net motion, even in the presence of a large external load.[1–3] From an energetics point of view, RNAP is a molecular motor capable of converting chemical energy derived from NTP hydrolysis into mechanical work.[4] Unlike other molecular motors such as kinesin or myosin that move along a uniform track, RNAP transcribes on a DNA substrate with varying sequence content. This variability of sequences significantly affects the kinetics of RNAP motion and results in non-uniform elongation rates.[5,6] Theoretical modeling of transcription has so far focused on two main aspects of elongation: (1) how RNAP couples chemical catalysis energy to its translocation and mechanical work and (2) how its motion is regulated by the DNA sequence.

RSC Biomolecular Sciences
RNA Polymerases as Molecular Motors
Edited by Henri Buc and Terence Strick

Published by the Royal Society of Chemistry, www.rsc.org

The first question concerns the mechano-chemical coupling mechanism of RNAP and applies to any molecular motor, or more generally, to any motor. However, there is a major difference between molecular and macroscopic motors: molecular motors function on such a small scale that they are significantly affected by thermal (Brownian) fluctuations. Because Brownian motion is random, it alone cannot generate unidirectional motor motion; nevertheless, it plays an important role in the functioning of molecular motors.[7] Two distinct mechano-chemical coupling mechanisms have been proposed that differ in the way the motor utilizes thermal and chemical energies. In the "power-stroke" mechanism, the energy derived from chemical reaction is used directly to drive the motor forward. In the alternative "Brownian ratchet" mechanism, the motor's motion is driven by thermal fluctuations, while the chemical reaction imposes directionality by biasing the motion in a single direction.

The mechano-chemical coupling mechanism employed by RNAP is still under debate (see Chapters 4 and 7). The power-stroke mechanism was proposed based on crystallographic studies of T7 RNAP[8] and suggests that the release of PPi product at the end of the chemical reaction step induces a conformational change in RNAP that makes it forward translocate by 1 bp. However, several other experimental and theoretical studies[9–17] supported a Brownian ratchet mechanism where RNAP can slide back and forth on the DNA template activated by thermal energy and the incorporation of the next nucleotide biases the polymerase forward by one base pair. The most direct evidence for a Brownian ratchet model is the ability of the RNAP to non-catalytically slide backward along the DNA (backtrack) when elongation is blocked *via* NTP starvation or slowed at particular DNA sequences.[11] However, it is not clear whether such thermally-activated sliding occurs at all template positions during active elongation.[18] Recent kinetic modeling combined with single-molecule measurements of RNAP elongation kinetics under various experimental conditions have provided strong support for a Brownian ratchet mechanism.[12,13,15,17]

The second key question that theoretical modeling of transcription must address is the influence of the underlying DNA sequence on RNAP kinetics. Experimentally, it has long been recognized that on a DNA template with varying ATGC content the elongation reaction does not proceed at a uniform rate. In particular, RNAP tends to dwell at some template positions considerably longer than at others, a phenomenon known as transcriptional pausing.[5,6] Bulk biochemical assays have suggested that at least some of the observed pauses were caused by RNAP backtracking.[6,11] At such pause sites transcription is halted until the 3′ end of the nascent RNA returns to the active site, either by RNAP forward translocation or by internal cleavage of the RNA at the active site. Single-molecule studies found that pauses could be divided into long/short duration pauses and that the two types of pauses exhibited different sensitivity to external force.[3,19,20] Measurements with improved spatial resolution revealed that RNAP tended to backtrack during the long pauses, but not the short ones, suggesting different pause mechanisms.[21,22] The DNA

sequences that are known to induce pauses lack apparent sequence consensus. Theoretical studies suggest that the sequence-dependence of RNAP elongation kinetics is highly correlated with the free energy of the corresponding transcription elongation complex (TEC), which depends strongly on the underlying DNA sequence.[12,17,23,24] These models have been successful in predicting a large portion of experimentally identified pause sites, and provided insight into the mechanism of pausing.

9.2 Background

RNAPs occur as both single- and multiple-subunit enzymes. RNAPs from bacteriophages and mitochondria are representative of the single-subunit family; bacterial, archaeal, and eukaryotic nuclear RNAPs constitute the multiple-subunit family. Although the single-subunit and multi-subunit RNAPs do not likely share a common ancestor, the available biochemical and structural information from representatives of each family shows that these RNAPs share many characteristics,[25,26] a justification for the search of common mechanisms.

Transcription is traditionally divided into three sequential phases: initiation, elongation and termination, although termination can be viewed as an alternate pathway branching from elongation.[27] Here we focus on the elongation stage, during which the RNAP forms a stable transcription elongation complex (TEC) with the template DNA and the nascent RNA, and moves along the DNA incorporating complementary NTPs to the 3′ end of the RNA. Each nucleotide addition can be viewed as a competition among active elongation, pausing (a transient conformational state incapable of elongation), arrest (a conformational state incapable of elongation without factor-assisted isomerization back to an active complex) and termination (transcript release and enzyme dissociation from the DNA template) (see Chapters 7 and 10 in the present book).[27]

As depicted in the Structural Atlas and reviewed in Chapter 7, the transcribing RNAP has a crab claw shape with the "jaws" surrounding a central channel, which holds the nucleic acids. The TEC structure contains an open DNA bubble of 12–14 bp and an 8–9 bp RNA-DNA hybrid within the RNAP main internal channel, a smaller secondary channel or pore that likely serves as an entry channel for NTPs, and an RNA exit channel.[26,28,29] The active site, which is the catalytic center of the complex, is located at the junction of the main channel and secondary channels and contains at least one nucleotide binding site and a tightly bound Mg^{2+}. RNAP is a two-metal ion-dependent enzyme and the second active-site Mg^{2+} is thought to be coordinated with the incoming NTP (Chapters 7 and 8).[30]

For a given nascent RNA length, the TEC may exist in slightly different configurations that are due to RNAP translocation, which we will refer to as "translocation states" (see Figure 7.2). The most significant difference among different translocation states is the position of the RNA 3′ end relative to the active site of RNAP. NTP incorporation can only occur in the

"post-translocation state," where the active site is occupied by the complementary NTP and immediately adjacent to the 3′ end of the nascent RNA. Upon NTP incorporation, the 3′ end of the RNA occupies the active site and the TEC is in its "pre-translocation state." From this state RNAP must move forward by 1 bp to clear the active site to allow binding by the next incoming NTP and thus to start the next elongation cycle. As mentioned before, it was observed experimentally that TEC can also form "backtracked states" by moving backward non-catalytically and placing the 3′ end of RNA into the secondary channel. There is also evidence suggesting that the RNAP can translocate forward beyond the pre-translocation state to "forward-tracked states," which could serve as precursors to transcription termination.[31,32] Such configurations are not catalytically competent and the RNAP must return to the post-translocation state to continue elongation. The pre- and post-translocation states are indispensable for the continuous NTP incorporation cycle, and therefore belong to the "main" elongation pathway; the backtracked and forward-tracked states are part of the "branch" non-essential pathways.

9.3 Mechano-chemical Coupling of Transcription

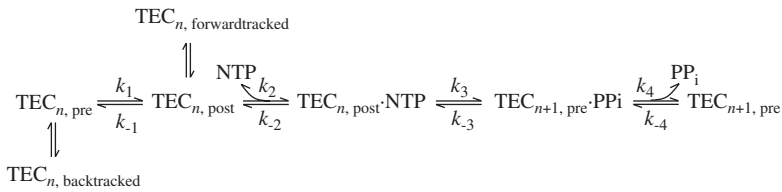
9.3.1 NTP Incorporation Cycle

A single NTP incorporation cycle must minimally include RNAP translocation from pre- to post-translocation states, NTP binding, NTP hydrolysis and PPi release.[18,33,34] However, the details of the NTP incorporation cycle depend greatly on the specific mechano-chemical coupling mechanism. In a Brownian ratchet model of transcription, prior to NTP incorporation, RNAP can move back and forth on the DNA template, activated solely by thermal energy, so the translocation part of the pathway is largely independent of the chemical reaction part of the pathway. In contrast, in a power-stroke model these two parts are highly correlated: NTP hydrolysis is thought to induce a conformational change of RNAP, which necessarily leads to translocation (as illustrated in Chapter 4 and Figure 7.5A). In other words, the two mechanisms represent "weak" and "strong" coupling between chemical reaction and mechanical translocation, respectively. Consequently, these models make different quantitative predictions of the elongation rate dependence on external load and [NTP], and can be differentiated by kinetic and mechanical measurements.

9.3.2 NTP Incorporation Pathway in a Simple Brownian Ratchet Model

The basic reaction pathway for a Brownian ratchet model involves only one NTP binding site and a minimum number of translocation states as

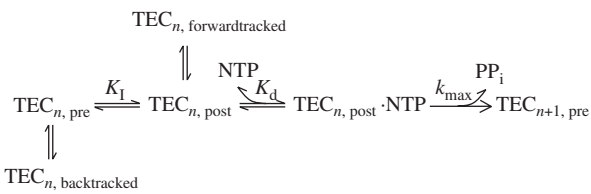
shown in:



(9.1)

where TEC$_{n,\text{pre(post)}}$ represents the TEC with transcript size n at the pre(post)-translocation state. The main reaction pathway proceeds along the horizontal arrows and includes translocation between pre- and post-translocation state, NTP binding, NTP hydrolysis and PPi release, all of which are potentially reversible.

Reaction pathway (9.1) could be further simplified to reaction pathway (9.2) below certain kinetic limits.[9,12] Under typical experimental conditions where PPi concentration in the solution is very low, the slow pyrophosphorolysis rate makes NTP hydrolysis essentially irreversible ($k_{-4} \sim 0$). Also, there is evidence that NTP hydrolysis rates are much larger than the PPi release rate ($k_{\pm 3} >> k_4$),[35] so the two steps could be combined into one step with a single effective rate. Another simplifying assumption is that both the translocation between the pre- and post- translocation states and NTP binding follow rapid equilibrium kinetics,[9] therefore only the equilibrium constants of these steps need to be considered:



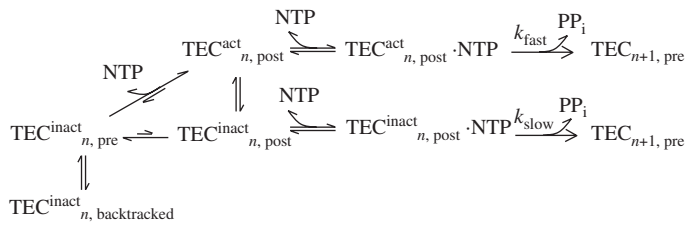
(9.2)

Notably, in this model, once a complementary NTP binds to the active site, the TEC is locked into the post-translocation state without access to other translocation states until NTP hydrolysis or dissociation. Therefore, it has been suggested that the NTP acts as a pawl in a ratchet and prevents the RNAP from sliding backwards,[14] consistent with the experimental observation that the incoming NTP stabilizes the TEC in its post-translocation state.[14]

9.3.3 NTP Incorporation Pathways in more Elaborate Brownian Ratchet Models

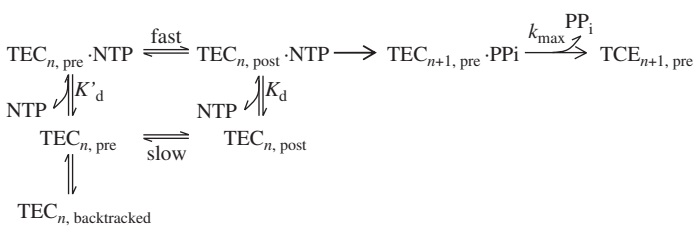
Brownian ratchet models more complex than the simple model described above have also been proposed and involve multiple parallel pathways,

additional TEC states, additional NTP binding sites, *etc*. In a quench flow experiment that probed *E. coli* RNAP transient-state kinetics, the rate of single nucleotide incorporation was found to be biphasic: a fraction of the RNAP population exhibited much faster rates of catalysis than the rest.[35,36] Therefore, it was proposed that the elongation complex could exist in either an active (TEC^{act}) or inactive (TEC^{inact}) state, which incorporate NTPs with different rates and are not in fast equilibrium, as shown in reaction pathway (9.3). Based on the dependence of the reaction rate on [NTP], it was also suggested that the transition from an inactive to an active state was facilitated by the binding of the next complementary NTP to an allosteric binding site:



(9.3)

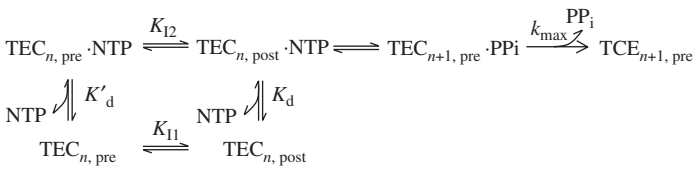
Multiphasic NTP incorporation kinetics were also proposed for human RNA polymerase II (pol II), although the kinetic data were modeled somewhat differently.[37,38] In the pol II model, the fast and slow fractions were proposed to reflect the TEC populations in the post- and pre-translocation states, respectively. The two translocation states interconvert slowly and are not in equilibrium, resulting in biphasic kinetics of NTP incorporation at low [NTP]. Because the two states respond to [NTP] differently, it was proposed that both translocation states could bind to the incoming NTP, and the transition between them was expedited by an NTP-driven "induced fit" mechanism:



(9.4)

A model similar to reaction path (9.4) was proposed in a recent single-molecule study that examined *E. coli* RNAP transcription rate dependence on [NTP] and force. This model also includes a secondary NTP binding site so that NTP binding can take place either before or after translocation.[13] However, unlike reaction path (9.4), the two translocation steps in this model, before and

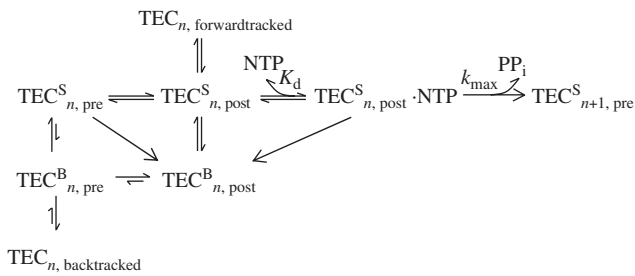
after NTP binding, are assumed to be in equilibrium:



(9.5)

Compared with the simple Brownian ratchet model (reaction path 9.2), models (9.3)–(9.5) all involve parallel pathways, with the transition between them depending on an NTP binding to a secondary site. This site was postulated to either lie within the streptolydigin-binding region of the *E. coli* RNAP or in the main channel facing the downstream single stranded DNA at $n+1$ and $n+2$ (where n is the RNA length).[36,38] However, so far direct experimental evidence for the extra NTP binding site is still lacking.

An alternative model proposed by Bar-Nahum *et al.*[14] involves parallel pathways, but with no requirement for an extra NTP binding site. This "dual ratchet mechanism" is based on experimental evidence that the bending of a flexible bridge helix (the F bridge) located at the active site near the 3′ end of the RNA plays an important role in translocation (see Figure 7.1.A).[39–41] According to this model, in addition to the simple Brownian ratchet mechanism in reaction path (9.2), there is a second ratchet imposed by the F bridge, which oscillates back and forth between "straight" and "bent" conformations due to thermal fluctuation. The F helix, the 3′ end of RNA and the incoming NTP substrate interact with each other by competing for the polymerase's active center: an NTP can only bind when the F helix is in its straight form, and the bending of the F helix can force a bound NTP to dissociate, to "push" RNAP forward by one nucleotide, or to cause the 3′ end of RNA to separate from the template DNA into the secondary channel. The reaction pathway is given by pathway (9.6):



(9.6)

where the superscripts S and B denote the straight and bent F bridge conformations, respectively.

9.3.4 NTP Incorporation Pathway in a Power-stroke Model

A power-stroke model has been proposed for the single subunit T7 RNA polymerase based on structural studies where snapshots of the TEC at different

stages of the elongation cycle were captured. T7 RNAP was found to undergo a large conformational change from an "open" conformation that could bind the incoming NTP to a "closed" conformation during catalysis (Chapter 4).[8,42] The transition from the "open" to "closed" state occurred upon binding of the NTP to the O helix structure located at the active site, while the pyrophosphate release reversed the transition with a concurrent forward translocation of RNAP by 1 bp. No structure could be obtained of the TEC in a pre-translocation state without the PPi bound within the active site. From these observations Yin and Steitz proposed a power-stroke mechanism in which the chemical energy derived from the NTP hydrolysis reaction directly drives the forward translocation of the RNAP along the DNA template. Because of the tight coupling of the translocation and the PPi release, they can be considered as a single step in the reaction pathway:

$$\mathrm{TEC}_{N,\,\mathrm{post}} \overset{\mathrm{NTP}}{\underset{K_\mathrm{d}}{\rightleftharpoons}} \mathrm{TEC}_{N,\,\mathrm{post}}\cdot\mathrm{NTP} \rightleftharpoons \mathrm{TEC}_{N+1,\,\mathrm{pre}}\cdot\mathrm{PPi} \xrightarrow{k_\mathrm{max}} \overset{\mathrm{PP_i}}{\mathrm{TEC}_{N+1,\,\mathrm{post}}} \tag{9.7}$$

In contrast to the Brownian ratchet models, the translocation step is irreversible and occurs only once per NTP incorporation cycle.

9.3.5 Elongation Kinetics

With the simple kinetic scheme in reaction path (9.2), the overall active elongation rate along the main pathway follows Michaelis–Menten kinetics in the presence of competitive inhibitor:[9,12]

$$k_\mathrm{main} = \frac{k_\mathrm{max}[\mathrm{NTP}]}{K'_\mathrm{d} + [\mathrm{NTP}]}, \quad \text{and} \quad K'_\mathrm{d} = K_\mathrm{d}(1 + K_\mathrm{I}) \tag{9.8}$$

where the effective dissociation constant is larger than the actual NTP dissociation constant K_d, reflecting the competition at the active site between the 3′ end of RNA and NTP. Similar arguments apply to the reaction path (9.5), where more inhibition states are present:

$$k_\mathrm{main} = \frac{k'_\mathrm{max}[\mathrm{NTP}]}{K'_\mathrm{d} + [\mathrm{NTP}]}, \quad \text{and} \quad K'_\mathrm{d} = K_\mathrm{d} \cdot \frac{1 + K_\mathrm{I1}}{1 + K_\mathrm{I2}} \tag{9.9}$$

Importantly, kinetic Equations (9.8) and (9.9) very much rely on the assumption of a rapid equilibrium existing between all states which interconvert during translocation. The importance of this assumption may be illustrated with a comparison of reaction pathways (9.4) and (9.5), which predict different kinetics despite their close resemblance. The model in reaction path (9.5) predicts single-exponential kinetics for the NTP incorporation reaction, while the overall elongation rate changes hyperbolically with varying [NTP]. In contrast, the model reaction path (9.4) predicts double-exponential incorporation kinetics and sigmoidal dependence of overall rate on [NTP].

The kinetics in reaction path models (9.3), (9.4) and (9.6) are more complicated. Despite the apparent difference between reaction paths (9.3) and (9.4), the NTP incorporation kinetics predicted by the two models are similar. In both models the TEC population equilibrates between the pre- and post-translocation states during NTP starvation. Upon NTP addition, the TEC population quickly re-distributes between the NTP bound and unbound states (NTP binding in these models is assumed to be fast), and the distribution depends on the NTP concentration. Different TEC states lead to different NTP incorporation rates, and the resulting kinetics are bi-phasic. Increasing [NTP] plays dual roles: it increases the NTP incorporation rate according to Michaelis–Menten kinetics, and it also shifts the TEC population into an "active" state displaying faster hydrolysis, and thus the overall incorporation rate has a sigmoidal dependence on [NTP]. In contrast, in model reaction path (9.6), the transition between the main and parallel pathways does not depend on [NTP]. Therefore, although the model could account for the biphasic incorporation rate by assuming a slow structural oscillation coupled to polymerization (for instance trigger loop and/or bridge helix oscillations), it could not explain the high sensitivity of the reaction rate to [NTP].

For the power-stroke model in reaction path (9.7), if we assume the NTP binding and hydrolysis reaction are both much faster than the PPi release step, and $[\mathrm{TEC}_{n+1,\mathrm{pre}} \cdot \mathrm{PPi}] = [\mathrm{TEC}_{n,\mathrm{post}} \cdot \mathrm{NTP}] \cdot K$, then the incorporation rate could be expressed as:

$$k = \frac{k'_{\max}[\mathrm{NTP}]}{K'_{\mathrm{d}} + [\mathrm{NTP}]}$$

where the effective NTP dissociation constant $K'_{\mathrm{d}} = K_{\mathrm{d}}/(1+K)$ and $k'_{\max} = k_{\max}[K/(1+K)]$ is the effective maximum incorporation rate.

9.3.6 Force-dependent Elongation Kinetics

Several single-molecule experiments were carried out to evaluate reaction path models (9.2), (9.5) and (9.7).[13,15,17] Compared with traditional ensemble studies, there are several advantages of single-molecule approaches for studying transcription elongation. First, single-molecule techniques, such as optical and magnetic tweezers, allow application of external force or torque to the TEC, thus selectively perturbing the reaction steps involving RNAP translocation. The nature of this step defines precisely how force affects the overall elongation rate and, therefore, the force-dependence of elongation kinetics is a very critical probe for the elongation mechanism. Second, as mentioned above, elongation has a complicated, multi-branched reaction pathway, and different parts of the pathway may respond differently to perturbations. For instance, active elongation kinetics could have a very different force-dependence than pausing kinetics. By monitoring the RNAP position in real time, especially with high spatial resolution, it is possible to decouple active elongation from pausing, so that their kinetics may be analyzed separately. Third, single-molecule techniques are suitable for measuring properties that are highly unsynchronized

among different RNAP molecules (such as pause duration), because they do not require ensemble averaging among a large number of molecules.

Differentiation among reaction path models (9.2), (9.5) and (9.7) is difficult using traditional bulk assays because the predicted kinetics are very similar (see above). However, they predict significantly different force-dependent elongation kinetics. The power-stroke model in reaction path (9.7) contains a set of reversible reaction steps followed by a force-dependent irreversible step, therefore the external force should not affect the NTP dissociation constant but simply the maximal rate constant:

$$k(F) = k(0) \cdot \exp(F\Delta/k_{\mathrm{B}}T) = \frac{k'_{\max}[\mathrm{NTP}]}{K'_{\mathrm{d}} + [\mathrm{NTP}]} \exp(F\Delta/k_{\mathrm{B}}T) \tag{9.10}$$

where $k(0)$ is the reaction rate at force 0 pN, F is the applied force and Δ is a characteristic distance representing the location of the activation barrier from the pre-translocation state ($0 \leq \Delta \leq 3.4\,\mathrm{nm}$), and $k_{\mathrm{B}}T$ is the thermal energy (Figure 9.1A). When $\Delta = 0\,\mathrm{bp}$, there would be no force-dependence on the predicted elongation rate; in other cases, the elongation rate would have the same exponential dependence on F at any NTP concentration. The actual change in the elongation rate at different forces, $k(F_1) - k(F_2)$, would be larger at higher [NTP], making it easier to be detected experimentally.

In the Brownian-ratchet model (reaction path 9.2), the force-affected translocation step is reversible and in rapid equilibrium. The force affects the equilibrium constant K_{I} and thus the effective NTP dissociation constant by tilting the energy landscape between the pre- and post-translocation states, resulting in an altered elongation rate (see Equation 9.8):

$$\begin{aligned} k(F) &= \frac{k_{\max}[\mathrm{NTP}]}{K_{\mathrm{d}}[1 + K_{\mathrm{I}}(F)] + [\mathrm{NTP}]} \\ &= \frac{k_{\max}}{1 + \frac{K_{\mathrm{d}}}{[\mathrm{NTP}]}[1 + K_{\mathrm{I}}(0) \cdot \exp(-Fd/k_{\mathrm{B}}T)]} \end{aligned} \tag{9.11}$$

where the characteristic distance d corresponds to 1 bp for transitions between pre- and post-translocation states (Figure 9.1B). Contrary to the previous case, here when [NTP] is infinitely high, the maximal elongation rate would no longer depend on force and the force modulates the effective K_{d}. Therefore, the sensitivity of the elongation rate to force decreases at higher [NTP]. A similar derivation can be applied to reaction path model (9.5), where applied force shifts the equilibrium distribution between different translocation states. Because in this model translocation could occur either before or after the NTP binding, the force effect contains both NTP-dependent and independent components:

$$k(F) = \frac{k_{\max}}{1 + K_{\mathrm{I2}}(0) \cdot \exp(-Fd/k_{\mathrm{B}}T) + \frac{K_{\mathrm{d}}}{[\mathrm{NTP}]}[1 + K_{\mathrm{I1}}(0) \cdot \exp(-Fd/k_{\mathrm{B}}T)]} \tag{9.12}$$

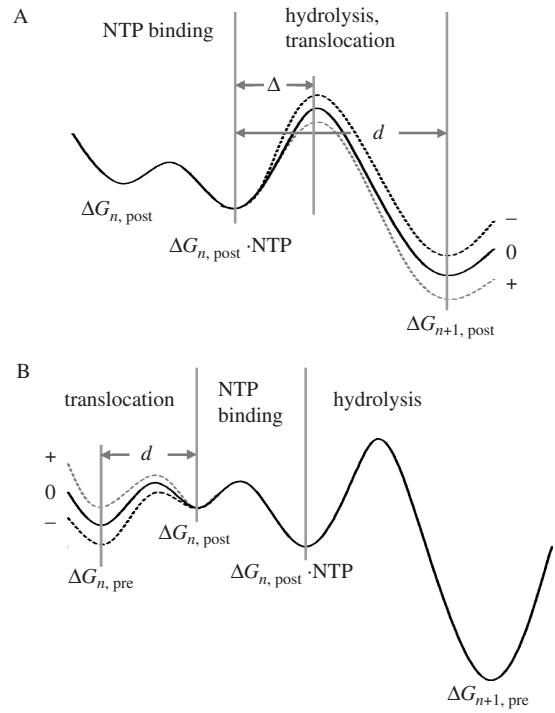


Figure 9.1 Cartoons illustrating the effect of an applied force on the energy landscape corresponding to addition of a single nucleotide during elongation. (A) Power-stroke model. (B) Brownian ratchet model. In both cases, force selectively perturbs the energy landscape in the step involving translocation. Assisting (+) and opposing (−) forces tilt the landscape in opposite directions relative to the zero force case. See text for details.

Therefore, reaction path model (9.5) also predicts a decreasing force-dependent elongation velocity at higher [NTP], but the force-dependence approaches a non-zero limit at very high [NTP].

The Brownian ratchet and power-stroke models predict different trends of elongation velocity force-dependence *vs.* [NTP]; therefore, single-molecule force–velocity measurements can be used to differentiate between the models. It has been shown that *E. coli* RNAP is a powerful motor, and on average ~25 pN of resisting force was required to stall elongation.[1–3] Under saturating [NTP], elongation velocity remained nearly constant over a wide range of resisting and assisting forces (+25 to −35 pN).[2,3,20,43] Two more recent studies on *E. coli* RNAP found a small but statistically significant dependence of velocity on force, even at high NTP concentrations, and this dependence increased at lower [NTP].[13,17] A study using T7 RNAP found a similar force–velocity relation, arguing for a similar elongation mechanism for the single and multiple subunit polymerases.[15]

Importantly, in reaction path model (9.2), the equilibrium translocation constant K_I can be computed solely based on the sequence-dependent TEC state energy, as will be discussed in detail below. Bai *et al.* showed that once k_{max} and K_d were experimentally determined, the force–velocity curves at various NTP concentrations in reaction path model (9.2) could be predicted without any fitting parameters (Equation 9.11).[17] The agreement between the predicted and the measured force–velocity curves supports the simple Brownian-ratchet mechanism (reaction path 9.2).[17]

Because the power-stroke model as presented in reaction path (9.7) is oversimplified and contains multiple assumptions on reaction kinetics, the power-stroke mechanism cannot be formally ruled out. However, generally speaking, the Brownian ratchet models have found strong support from independent experimental evidence despite some differences in the details of the reaction pathways. In Thomen *et al.*[15] and Bai *et al.*[17] the measured force–velocity relationships of T7 and *E. coli* RNAP at various NTP concentrations were consistent with the simple Brownian ratchet model of reaction path (9.2), while Abbondanzieri *et al.*[13] found that their data could be better explained by the more complicated Brownian ratchet model reaction path (9.5).

9.4 Sequence-dependent RNAP Kinetics

To model sequence-dependent RNAP kinetics, the essential task is to establish the correlation between the kinetic parameters in the NTP incorporation pathway and the sequence of the transcribed DNA. A hint comes from studies of intrinsic transcription termination, where a special sequence known as a terminator destabilizes the TEC and induces dissociation of the RNA and RNAP from the DNA template (Chapter 10).[44] To account for this destabilization, Yager and von Hippel[23] provided a static sequence-dependent thermodynamic analysis of TEC stability of *E. coli* RNAP. This model successfully explained the elevated TEC free energy at intrinsic terminators, and laid out the foundation for kinetic analysis of termination efficiency.[45] Recent theoretical studies extended the thermodynamic model by performing a full kinetic analysis and applying it to active elongation and pausing.[12,17,24] These kinetic models are discussed below.

9.4.1 Thermodynamic Analysis of the TEC

Based on the structure of the TEC, its thermodynamic stability is given by the standard free energy of the complex formation from its components:

$$\Delta G = \Delta G_{\text{DNA bubble}} + \Delta G_{\text{RNA–DNA hybrid}} + \Delta G_{\text{RNAP binding}} \tag{9.13}$$

Here the lower the free energy of a given state the more stable the TEC in that state. The energy cost required for DNA bubble opening $\Delta G_{\text{DNA bubble}}$ is offset by the energy of formation of the RNA-DNA hybrid $\Delta G_{\text{RNA–DNA hybrid}}$ and

the stabilizing interactions between the polymerase and the nucleic acids, $\Delta G_{\text{RNA binding}}$. The first two terms of Equation (9.13) can be calculated for a given DNA sequence from the base pairing energy of the nucleic acids. These free energies have been measured in solution only and current models assume the energy values do not change considerably within the TEC. Based on structural data, RNAP interacts primarily with the backbone of the nucleic acids and thus the third term in Equation (9.13) can be treated as a sequence-independent constant.[23]

Another possible energetic contribution that may affect RNAP kinetics comes from the interactions of RNAP with the nascent RNA, ΔG_{RNA}:

$$\Delta G = \Delta G_{\text{DNA bubble}} + \Delta G_{\text{RNA–DNA hybrid}} + \Delta G_{\text{RNAP binding}} + \Delta G_{\text{RNA}} \qquad (9.14)$$

The last term represents the change in the folding energy of the free transcript RNA outside the RNAP as transcription proceeds.[24] RNA is known to play a critical role in intrinsic transcription termination and hairpin-dependent transcriptional pausing.[6,46] It has also been proposed that RNA hairpin structures could serve as a barrier to prevent RNAP backtracking and thus influence elongation pausing.[24]

The value of the TEC state energy is highly dependent on the TEC structure used in modeling elongation (size of the DNA bubble, length of the DNA-RNA hybrid, *etc.*). Recent theoretical studies computed the TEC energy either using a fixed TEC structure within the experimentally observed range[12,17,23] or averaging over a range of bubble configurations.[24] For a given TEC structure, the energy depends strongly on the DNA sequence within the complex. On a DNA template of a few hundred basepairs, the TEC energy at different template positions could differ by more than $10k_BT$. This large dynamic range has important implications and naturally leads to the dramatic sequence-dependent kinetic behavior of RNAP.

This overall free energy also depends on the precise translocation state of the TEC. The post-translocation and forward-tracked states tend to be less energetically favorable than the pre-translocation and backtracked states, because the former have fewer base pairs within the RNA-DNA hybrid. Based on the same line of reasoning, the pre-translocation and backtracked states would have approximately the same energy, once averaged over various DNA sequences. Once an incoming complementary NTP binds in the active site and basepairs with the DNA base, this additional interaction energy stabilizes the post-translocation configuration.[8,9]

9.4.2 Sequence-dependent NTP Incorporation Kinetics in Brownian Ratchet Models

Thus far two sequence-dependent elongation models have been formulated, respectively by Bai *et al.*[12,17] and by Tadigotla *et al.*[24] Both models are based on the simple Brownian ratchet mechanism shown in kinetic pathway (9.1).

They are extended to consider sequence-dependent k_{max}, K_d and translocation rates in the branched pathways shown in reaction path (9.2) for all template positions. Since there have been little or no direct measurements of these kinetic parameters, some assumptions must be made to estimate the parameter values. The key differences between the two models are in the assumptions leading to the derivation of the backtracking rates. They lead to different expressions for the elongation rate.

In Bai *et al.*[12,17] backtracking is assumed to be a slow process, so that during active elongation the backtracked states do not equilibrate with the states within the main pathway. At a majority of template positions and under typical NTP concentrations this model predicts that the probabilities of backtracking are so small that the backtracked states are essentially inaccessible. For these sites, the NTP incorporation follows a single exponential curve with a rate k_{main} as expressed in Equation (9.8). Notably, the equilibrium constant, K_I, is not a free parameter and can be directly calculated from the free energy difference between the pre- and post-translocation states:

$$K_I = \exp\left[\left(\Delta G_{post} - \Delta G_{pre}\right)/k_B T\right] \tag{9.15}$$

As mentioned before, ΔG_{pre} and ΔG_{post} are both strongly sequence-dependent, and thus K'_d as well as k_{main} will be sequence-dependent.

In this model the backtracked rates are calculated based on Arrhenius kinetics, where the transition rate between any two states 1 and 2 is governed by the free energy difference between the peak of the activation barrier for the transition and state 1:

$$k_{1\to 2} = k_0 \exp[-(\Delta G^{\dagger}_{1\leftrightarrow 2} - \Delta G_1)/k_B T] \tag{9.16}$$

where $\Delta G^{\dagger}{}_{1\leftrightarrow 2}$ is assumed to be the same for all backtracked steps and on all template positions. Under such an assumption, the backtracking rate is solely determined by the TEC free energies. At the few sites displaying a high probability of backtracking, NTP incorporation is predicted to follow a multiphasic time course.

In contrast, Tadigotla *et al.*[24] assume that RNAP is capable of rapid backtracking until it encounters the first secondary structure formed by the nascent RNA outside of the RNAP. At this point, further backtracking of RNAP is prevented until the next NTP incorporation. In this model, accessible translocation states both in the main and branch pathways are assumed to be in equilibrium. Thus at each template position the predicted time course of NTP incorporation follows a single exponential with a rate constant:

$$k_{main} = \frac{k_{max}[NTP]}{K'_d + [NTP]} = \frac{k_{max}[NTP]}{K_d \sum_m \exp\left(\Delta G_{post} - \Delta G_m\right)/k_B T + [NTP]} \tag{9.17}$$

which depends on the free energy of all accessible states, ΔG_m.

9.4.3 Model Predictions of Pause Locations, Kinetics and Mechanisms

The models of sequence-dependent elongation kinetics described above may be used to predict pause locations and pause kinetics, as well as to provide insights into pause mechanisms. Interestingly, these two models do not always make the same predictions for pausing.

The model in Bai *et al*. predicts two alternative pausing mechanisms: pausing could be caused either by a slow rate within the main pathway (''on-pathway'' pause) or by a relatively fast rate of entry into a nonproductive branch pathway with a slow rate of returning to the main pathway (backtracking pause). Note that the two types of pauses are correlated because different pathways are kinetically competitive: a slow rate in the main pathway increases the probability for the RNAP to enter a branched pathway. Equation (9.15) shows that an on-pathway pause occurs at a template site with a large $\Delta G_{\text{post}} - \Delta G_{\text{pre}}$, *i.e*., the pre-translocation state is much more stable than the post-translocation state. Thus the entire RNAP population is expected to follow the same slow kinetics. We refer to this type of pause as a pre-translocation pause. In contrast, a backtracking pause is expected to occur when the pre-translocation state is unstable (high ΔG_{pre}) so that $k_{\text{pre}\rightarrow\text{backtracking}}$ becomes significant in comparison with k_{main} (Equation 9.8). The dwell time distribution at backtracking pause sites is no longer defined by a single characteristic time but has a long tail due to the slow return of the backtracked RNAP population to the main pathway. Compared to a pre-translocation pause, a backtracking pause tends to occur less frequently but for a longer duration. Because backtracking typically involves translocation over several base pairs (as opposed to 1 bp in the on-pathway pause), a backtracking pause is more sensitive to external force. The two types of pauses therefore provide an explanation for the short and long duration pauses observed in single-molecule experiments.[21,22] This model also correctly predicted prominent pause positions seen in bulk studies, and the force-dependence of pause durations for a known pause sequence detected by single-molecule experiments.[17,47]

In the model of Tadigotla *et al*., pauses occur when one or more accessible translocation states are significantly more stable than the post-translocation state (Equation 9.17). This model was shown to have successfully predicted a large portion of previously identified pauses.[5] Because most of the pause sites in Bai *et al*. also have unstable post-translocation states, the two models would have significant overlap in pause sites prediction. However, the pausing kinetics and mechanism of the pauses predicted by the two models are different: most of the pauses predicted by Tadigotla *et al*. are backtracked, and the NTP incorporation kinetics at these sites all follow a single exponential. Future effort is required to differentiate and evaluate the two models by comparison with more experimental data.

The models mentioned above are mostly based on transcription studies *in vitro*, but have direct implications for cellular processes *in vivo*. It is known that transcription levels of certain genes can exhibit high cell to cell

variation (noise), especially when the transcripts are generated in a burst-like fashion.[48–53] Usually, the noise is thought to originate from infrequent transcription initiation.[51–53] A recent model demonstrated that long transcriptional pausing could also lead to bursts of mRNA production and may be a significant contributor to the variability in cellular transcription rates.[54] Further experimental and modeling studies of transcriptional rate, pausing and interactions between elongating RNAPs would shed light on both the fundamental elongation mechanism and the corresponding physiological effects on cell growth and viability.

Acknowledgements

This work is supported by grants to M.D.W. from NSF grant (DMR-0517349), NIH grant (R01 GM059849), and the Keck Foundation's Distinguished Young Scholar Award, as well as postdoctoral fellowships to L.B. from the Rockefeller University and the Damon Runyon Cancer Research Foundation.

References

1. H. Yin, M. Wang, K. Svoboda, R. Landick, S. Block and J. Gelles, *Science*, 1995, **270**, 1653–1657.
2. M. Wang, M. Schnitzer, H. Yin, R. Landick, J. Gelles and S. Block, *Science*, 1998, **282**, 902–907.
3. K. Neuman, E. Abbondanzieri, R. Landick, J. Gelles and S. Block, *Cell*, 2003, **115**, 437–447.
4. J. Gelles and R. Landick, *Cell*, 1998, **93**, 13–16.
5. J. Levin and M. Chamberlin, *J. Mol. Biol.*, 1987, **196**, 61–84.
6. I. Artsimovitch and R. Landick, *Proc. Natl. Acad. Sci. USA*, 2000, **97**, 7090–7095.
7. R. Feynman, R. Leighton and M. Sands, *in The Feynman Lectures on Physics*, Addison-Wesley Pub. Co., Reading, Mass, 1963.
8. Y. Yin and T. Steitz, *Cell*, 2004, **116**, 393–404.
9. R. Guajardo and R. Sousa, *J. Mol. Biol.*, 1997, **265**, 8–19.
10. F. Jülicher and R. Bruinsma, *Biophys. J.*, 1998, **74**, 1169–1185.
11. N. Komissarova and M. Kashlev, *J. Biol. Chem.*, 1997, **272**, 15329–15338.
12. L. Bai, A. Shundrovsky and M. Wang, *J. Mol. Biol.*, 2004, **344**, 335–349.
13. E. Abbondanzieri, W. Greenleaf, J. Shaevitz, R. Landick and S. Block, *Nature*, 2005, **438**, 460–465.
14. G. Bar-Nahum, V. Epshtein, A. Ruckenstein, R. Rafikov, A. Mustaev and E. Nudler, *Cell*, 2005, **120**, 183–193.
15. P. Thomen, P. Lopez and F. Heslot, *Phys. Rev. Lett.*, 2005, **94**, 128102.
16. Q. Guo and R. Sousa, *J. Mol. Biol.*, 2006, **358**, 241–254.
17. L. Bai, R. Fulbright and M. Wang, *Phys. Rev. Lett.*, 2007, **98**, 068103.
18. R. Landick, *Cell*, 2004, **116**, 351–353.

19. K. Adelman, A. La Porta, T. Santangelo, J. Lis, J. Roberts and M. Wang, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 13538–13543.
20. N. Forde, D. Izhaky, G. Woodcock, G. Wuite and C. Bustamante, *Proc. Natl. Acad. Sci. USA*, 2002, **99**, 11682–11687.
21. J. Shaevitz, E. Abbondanzieri, R. Landick and S. Block, *Nature*, 2003, **426**, 684–687.
22. K. Herbert, A. La Porta, B. Wong, R. Mooney, K. Neuman, R. Landick and S. Block, *Cell*, 2006, **125**, 1083–1094.
23. T. Yager and P. von Hippel, *Biochemistry*, 1991, **30**, 1097–1118.
24. V. Tadigotla, D. O Maoiléidigh, A. Sengupta, V. Epshtein, R. Ebright, E. Nudler and A. Ruckenstein, *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 4439–4444.
25. R. Sousa, *Trends. Biochem. Sci.*, 1996, **21**, 186–90.
26. P. Cramer, *Bioessays*, 2002, **24**, 724–729.
27. P. von Hippel, *Science*, 1998, **281**, 660–665.
28. E. Nudler, I. Gusarov, E. Avetissova, M. Kozlov and A. Goldfarb, *Science*, 1998, **281**, 424–428.
29. S. Darst, *Curr. Opin. Struct. Biol.*, 2001, **11**, 155–162.
30. V. Sosunov, E. Sosunova, A. Mustaev, I. Bass, V. Nikiforov and A. Goldfarb, *EMBO J.*, 2003, **22**, 2234–2244.
31. W. Yarnell and J. Roberts, *Science*, 1999, **284**, 611–615.
32. T. Santangelo and J. Roberts, *Mol. Cell*, 2004, **14**, 117–126.
33. D. Erie, T. Yager and P. von Hippel, *Annu. Rev. Biophys. Biomol. Struct.*, 1992, **21**, 379–415.
34. S. Uptain, C. Kane and M. Chamberlin, *Annu. Rev. Biochem.*, 1997, **66**, 117–172.
35. J. Foster, S. Holmes and D. Erie, *Cell*, 2001, **106**, 243–252.
36. S. Holmes and D. Erie, *J. Biol. Chem.*, 2003, **278**, 35597–35608.
37. Y. Nedialkov, X. Gong, S. Hovde, Y. Yamaguchi, H. Handa, J. Geiger, H. Yan and Z. Burton, *J. Biol. Chem.*, 2003, **278**, 18303–18312.
38. C. Zhang and Z. Burton, *J. Mol. Biol.*, 2004, **342**, 1085–1099.
39. A. Gnatt, P. Cramer, J. Fu, D. Bushnell and R. Kornberg, *Science*, 2001, **292**, 1876–82.
40. D. Vassylyev, S. Sekine, O. Laptenko, J. Lee, M. Vassylyeva, S. Borukhov and S. Yokoyama, *Nature*, 2002, **417**, 712–719.
41. V. Epshtein, A. Mustaev, V. Markovtsov, O. Bereshchenko, V. Nikiforov and A. Goldfarb, *Mol. Cell*, 2002, **10**, 623–634.
42. D. Temiakov, V. Patlan, M. Anaikin, W. McAllister, S. Yokoyama and D. Vassylyev, *Cell*, 2004, **116**, 381–391.
43. R. Davenport, G. Wuite, R. Landick and C. Bustamante, *Science*, 2000, **287**, 2497–2500.
44. K. Wilson and P. von Hippel, *J. Mol. Biol.*, 1994, **244**, 36–51.
45. P. von Hippel and T. Yager, *Proc. Natl. Acad. Sci. USA*, 1991, **88** 2307–2311.
46. I. Toulokhonov and R. Landick, *Mol. Cell*, 2003, **12**, 1125–1136.

47. A. Shundrovsky, T. Santangelo, J. Roberts and M. Wang, *Biophys. J.*, 2004, **87**, 3945–3953.
48. M. Elowitz, A. Levine, E. Siggia and P. Swain, *Science*, 2002, **297** 1183–1186.
49. E. Ozbudak, M. Thattai, I. Kurtser, A. Grossman and A. van Oudenaarden, *Nat. Genet.*, 2002, **31**, 69–73.
50. W. Blake, M. Kaern, C. Cantor and J. Collins, *Nature*, 2003, **422**, 633–637.
51. L. Cai, N. Friedman and X. Xie, *Nature*, 2006, **440**, 358–362.
52. J. Raser and E. O'Shea, *Science*, 2005, **309**, 2010–2013.
53. J. Yu, J. Xiao, X. Ren, K. Lao and X. Xie, *Science*, 2006, **311**, 1600–1603.
54. M. Voliotis, N. Cohen, C. Molina-Paris and T. Liverpool, *Biophys J.*, 2008, **94**, 334–348.