

## Force Generation in RNA Polymerase

Hong-Yun Wang,\* Tim Elston,\* Alexander Mogilner,# and George Oster\*

\*Department of Molecular and Cellular Biology, University of California, Berkeley, California 94720-3112, and #Department of Mathematics, University of California, Davis, California 95616 USA

**ABSTRACT** RNA polymerase (RNAP) is a processive molecular motor capable of generating forces of 25–30 pN, far in excess of any other known ATPase. This force derives from the hydrolysis free energy of nucleotides as they are incorporated into the growing RNA chain. The velocity of procession is limited by the rate of pyrophosphate release. Here we demonstrate how nucleotide triphosphate binding free energy can rectify the diffusion of RNAP, and show that this is sufficient to account for the quantitative features of the measured load-velocity curve. Predictions are made for the effect of changing pyrophosphate and nucleotide concentrations and for the statistical behavior of the system.

### GLOSSARY

$f$	load force from the laser trap (pN)
$k_B T$	4.1 pN-nm
$\alpha_1$	polymerization rate of RNA by the nucleotides hydrolyzed on RNAP (1/s)
$\alpha_2$	dissociation rate of $PP_i$ (1/s)
$\alpha_3$	polymerization rate of RNA by the nucleotides hydrolyzed in solution when a $PP_i$ is bound to the RNAP (1/s)
$\alpha_4$	polymerization rate of RNA by the nucleotides hydrolyzed in solution when no $PP_i$ is bound to the RNAP (1/s)
$\beta_1$	depolymerization rate of RNA (the reverse of $\alpha_1$ ) (1/s)
$\beta_2$	association rate of $PP_i$ (1/s)
$\beta_3$	depolymerization rate of RNA (the reverse of $\alpha_3$ ) (1/s)
$\beta_4$	depolymerization rate of RNA (the reverse of $\alpha_4$ ) (1/s)
$\delta$	length of a base pair (0.34 nm)
$\rho_1(n, t)$	the probability density that the RNAP is in state $P_n$ at time $t$ (i.e., with no $PP_i$ bound to the RNAP)
$\rho_2(n, t)$	the probability density that the RNAP is in state $P'_n$ at time $t$ (i.e., a $PP_i$ is bound on the RNAP)
$n$	Length of the RNA transcript

### INTRODUCTION

RNA polymerase (RNAP) processes along a DNA strand and transcribes the information coded in the DNA's base pair sequence into RNA (Erie et al., 1992; Polyakov et al., 1995; Yager and von Hippel, 1987). During this procession, RNAP polymerizes an RNA transcript via a sequence of

reactions taking place on its surface. First the incoming DNA chain is separated into single strands, one of which will serve as the template for the RNA. The two strands reanneal before they exit the posterior end of the enzyme; in between they form a "transcription bubble"  $\sim 15$  bp (1 bp = 0.34 nm) long (Fig. 1 *a*). The entire enzyme is  $\sim 30 \pm 5$  bp long. The growing RNA chain is synthesized at a catalytic site within the transcription bubble by the addition of nucleotides complementary to the sequence of the template strand. Viewed as a molecular machine, RNAP processes along the DNA by converting the free energy of nucleotide binding and hydrolysis into a force directed along the DNA axis. However, the molecular mechanism by which these chemical bond energies are transduced into the force that drives the RNAP along the DNA strand remains a mystery.

The velocity and force of procession can be measured by using laser trap technology to produce a force-velocity curve (M. D. Wang, personal communication). Measurements on RNA polymerase show that its mechanical properties differ from those measured for other molecular motors in two important respects. The load-velocity curve is concave, rather than convex or linear, as characteristic of other molecular motors (Berg and Turner, 1993; Finer et al., 1994; Hunt et al., 1994; Molloy et al., 1995; Svoboda and Block, 1994). Thus the velocity is nearly constant up to loads above 20 pN, whereupon it falls off to a stall load between 25 and 30 pN (M. D. Wang, personal communication). This stall force is 5–6 times larger than that of myosin or kinesin.

Here we propose a mechanism that can account for both of these features of the load-velocity curve, and which makes definite predictions about how the load-velocity curve changes when the concentrations of pyrophosphate and nucleotide are varied. The model also allows us to make predictions about the statistical behavior of the enzyme. In the next section we formulate the mathematical model. In the third section we present an analytic expression for the load-velocity curve, and show in the fourth section how the model parameters are computed from experimental data. In the fifth section we show how the load-velocity curve and the stall force should vary as pyrophosphate and nucleotide concentrations are changed. We also discuss how the model can be extended to incorporate DNA sequence-dependent

Received for publication 19 June 1997 and in final form 13 November 1997.

Address reprint requests to Dr. George Oster, Department of ESPN, University of California, 201 Wellman Hall, Berkeley, CA 94720-3112. Tel.: 510-642-5277; Fax: 510-642-5277; E-mail: goster@nature.berkeley.edu.

© 1998 by the Biophysical Society

0006-3495/98/03/1186/17 \$2.00

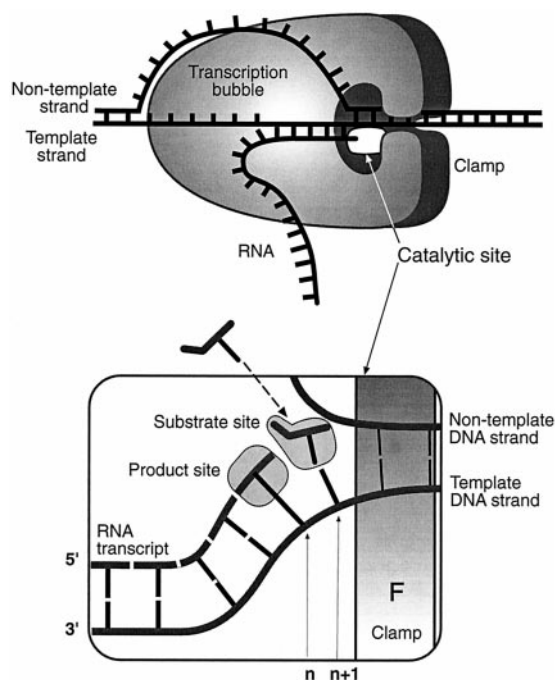


FIGURE 1 (a) Schematic diagram of polymerase. The *E. coli* RNAP consists of four subunits (denoted  $\alpha_2\beta\beta'$ ) with overall dimensions  $\sim 9 \times 11 \times 16$  nm. Approximately 30 bp fits into the major DNA groove (1 bp = 0.34 nm). During transcription, the entering DNA strand is separated into a template strand and a nontemplate strand to form a “transcription bubble” that is  $\sim 15$  bp long. Within this domain is the catalytic site where nucleotides are added to the growing RNA chain. The DNA-RNA hybrid region is 8–12 bp long. (b) The catalytic locus may consist of two binding sites (Erie et al., 1992). A substrate binding site on the surface of the enzyme binds solution NTP weakly. If the incoming base is complementary to the template base at the substrate site, they form a hydrogen bond. This triggers hydrolysis of the pyrophosphate group, and a phosphodiester link is forged with the preceding base (redrawn from Erie et al., 1992).

effects, such as pausing and backsliding (Landick, 1997), and the appearance of “inchworming” motion in DNA footprinting experiments (Chamberlin, 1994; Krummel and Chamberlin, 1992). We present only the key results in the body of the paper and place the mathematical derivations of these results in the Appendices.

## A MODEL FOR FORCE GENERATION IN RNAP

We shall discuss our model in the context of the experimental situation of Yin et al., shown schematically in Fig. 2 *a* (Yin et al., 1995; M. D. Wang, personal communication). In these experiments the RNAP is affixed to the substratum, and a 0.5- $\mu\text{m}$  bead attached to the end of the DNA is held in a laser trap. When nucleotide is added to the solution, the RNAP exerts tension on the DNA strand that pulls the bead off the center of the laser trap. Because the force developed by the laser trap is calibrated to piconewton accuracy, and the location of the bead can be tracked with nanometer precision, the force exerted on the bead by the RNAP can be accurately measured. In this fashion a load-velocity curve can be constructed. In practice, this is made difficult by the

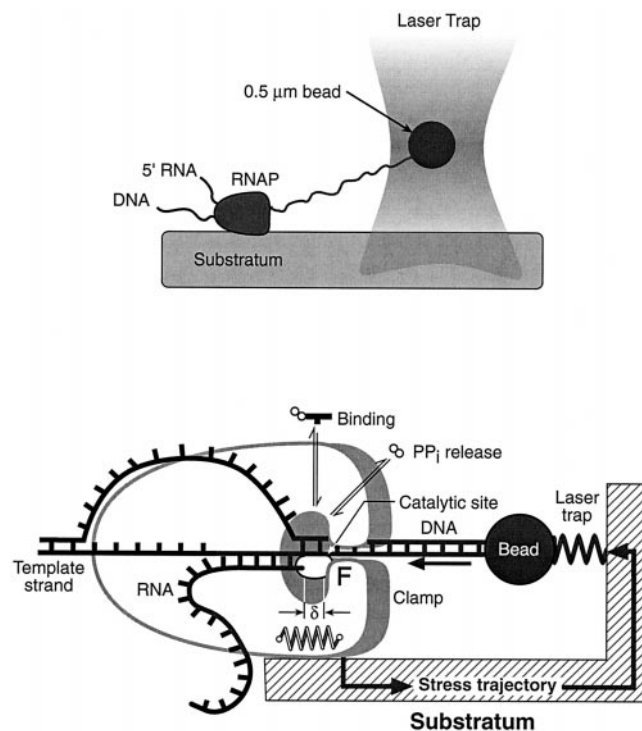


FIGURE 2 (a) Schematic diagram of the experimental setup (Yin et al., 1995). The RNAP is attached to a coverslip, and a 0.5- $\mu\text{m}$  bead is attached to the DNA strand. The position of the bead can be monitored optically. Procession of the RNAP reels in the DNA, pulling the bead out of the trap center. The trap is closely approximated by a linear elastic element, and so the force exerted on the DNA strand can be computed from the bead displacement. (b) Mechanical equivalent of the experimental setup used in formulating the model. Stress trajectories must be continuous; the stress flow is laser trap  $\rightarrow$  DNA  $\rightarrow$  RNA tip  $\rightarrow$  RNAP barrier  $\rightarrow$  substratum  $\rightarrow$  laser trap. The barrier *F* represents the interaction between the RNAP and the DNA-RNA hybrid, through which tension in the DNA strand is transferred to the RNAP, while the template DNA strand is allowed to pass freely. The clamp on the front part of RNAP is redrawn from Landick (1997). The protein elasticity is in series with the DNA and trap elasticities to form a composite elastic system. Intercalation of a hydrolyzed nucleotide between the growing tip of the RNA chain and the front end of the RNAP can occur if Brownian fluctuations in the system produce a gap larger than the size of a nucleotide. Until pyrophosphate is released, a new nucleotide cannot enter the RNAP binding site; thus  $\text{PP}_i$  release is the rate-limiting step for RNAP progression. The spring connecting the front and rear parts of the RNAP indicates that the two may move somewhat independently to produce the appearance of “inchworming” in DNA footprinting experiments.

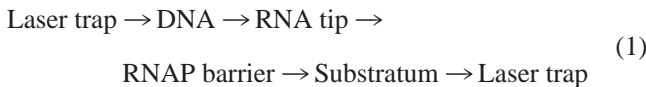
propensity of RNAP to pause intermittently in a sequence-dependent fashion.

## Mechanical assumptions

The geometry of the model is shown in Fig. 2 *b*. The DNA strand is attached at one end to the bead, which is held in the laser trap. The other end within the RNAP is annealed to the growing RNA strand via the 8–12-bp hybrid in the transcription bubble. The RNA strand, in turn, can make contact with the RNAP at a site at the front of the catalytic site (labeled *F* in Fig. 2 *b*). This site acts as a “barrier” against

which the tip of the transcript can fluctuate. In the model,  $F$  is the site where tension in the DNA strand is transferred to the RNAP, while allowing the template DNA strand to pass freely. Here the “barrier”  $F$  is meant to represent the “clamp” on RNAP (Landick, 1997), but it does not have to be a discrete site. Rather, it represents the overall interaction between the RNAP and the DNA-RNA hybrid, which prevents the DNA-RNA hybrid from being pulled out of the RNAP by the load force. The RNA transcript also may be attached to the RNAP at hybrid position posterior to the catalytic site and further back in an RNA before it exits the enzyme (Nudler et al., 1997). Both of these sites can transmit stress from DNA to RNAP. In the transcription process, RNAP needs to work against the load force and overcome the attachments of RNAP to DNA and RNA. It does this by rectifying the Brownian diffusion against the “barrier”  $F$  using the free energy of nucleotide triphosphate (NTP) binding and hydrolysis. Thus a dominant portion of the load force should flow through the “barrier”  $F$  to the substratum and then back to the laser trap.

Because stress trajectories must be continuous, we shall assume that the stress follows the path



We treat the enzyme as an elastic body attached firmly to the substratum. DNA, RNA, and laser traps are all modeled as springs. However, treating these objects as elastic bodies does not prevent us from focusing our attention only at the catalytic site. In Appendix B we demonstrate that the composite elastic forces can be very well approximated by a constant force, because the protein elasticity is in series with the DNA and trap elasticities.

We shall take advantage of the fact that there are three time scales inherent in the problem:

$t_M$  is the relaxation time to mechanical equilibrium between the RNA tip (the last subunit added) and the front of the catalytic site (labeled  $F$  in Fig. 2) after a polymerization event.  $t_M \approx \delta^2/2D_R$ , where  $D_R$  is the diffusion constant of the RNA tip relative to  $F$ , and  $\delta$  is the size of a nucleotide (0.34 nm).

$t_P$  is the time scale for NTP to enter the substrate site. This time scale depends on two coincidental processes: 1) diffusion of the NTP to the substrate site, and 2) the gap between the transcript tip and the barrier,  $F$ , being greater than  $\delta$ .

$t_R$  is the time scale of pyrophosphate release; this is the rate-limiting step in the progression of the RNAP.

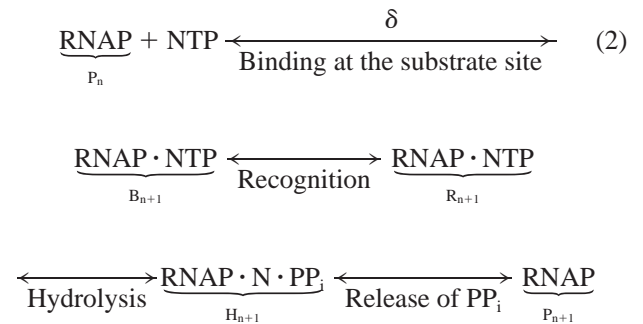
When  $t_M \ll t_P$ , i.e.,  $\delta/t_P \ll 2D_R/\delta$ , the system is always at thermodynamic equilibrium with respect to the tip of the transcript. (The time scale  $t_R$  does not disturb the relaxation to the thermodynamic equilibrium of the system. That is, even if  $PP_i$  release were very fast ( $t_R \ll t_M$ ), the system is still at thermodynamic equilibrium as long as the condition  $\delta/t_P \ll 2D/\delta$  is satisfied. Note that  $2D/\delta$  is the “perfect

ratchet velocity” (Peskin et al., 1993).) In Fig. 8 in Appendix A, we demonstrate that this condition is satisfied, even if we use the much smaller diffusion coefficient of the bead in the laser trap ( $\sim 10^5 \text{ nm}^2/\text{s}$ ). Therefore, we can safely assume an equilibrium distribution of the gap between the RNA tip and the RNAP “ratchet barrier,”  $F$ .

The key result of these calculations is that the rate of relaxing to the thermodynamic equilibrium after the addition of a nucleotide is much faster than the nucleotide insertion rate. So the distribution of the gap between the RNAP barrier and the transcript tip is time independent. This allows us to consider the motion of the RNAP to occur in discrete steps of length  $\delta = 0.34 \text{ nm}$ , the length of a single base pair. Thus we can formulate the model as a discrete state Markov chain.

### Kinetic assumptions

Transcript elongation takes place at the catalytic site shown schematically in Fig. 1 *b*. There is some ambiguity about the exact sequence of events taking place at the catalytic site (Erie et al., 1992). One possible sequence is



In this scheme the RNAP states are

$P_n$ : The transcript length is  $n$ .

$B_{n+1}$ : The transcript length is  $n + 1$  with NTP bound in the substrate site.

$R_{n+1}$ : The transcript length is  $n + 1$  with NTP bound in the substrate site and hydrogen bonded to the complementary site on the template DNA strand.

$H_{n+1}$ : The transcript length is  $n + 1$ , with the nucleotide bound in the substrate site and to the template DNA strand after the hydrolysis of nucleotide.

This scheme is illustrated graphically in Fig. 3 *a*, where we have plotted the spatial displacement,  $n$ , of the RNAP along the horizontal axis and the reaction coordinate along the vertical axis. All transition rates with a horizontal ( $n$ ) component depend on the load force,  $f$ , from the laser trap. We have placed the origin of our coordinate system at a particular nucleotide (e.g., the first) added to the growing RNA chain, and denoted by  $n$  the position of the RNA transcript tip. The subscript refers to the length of the transcript (or equivalently, the position of the RNAP from the beginning of the transcript).

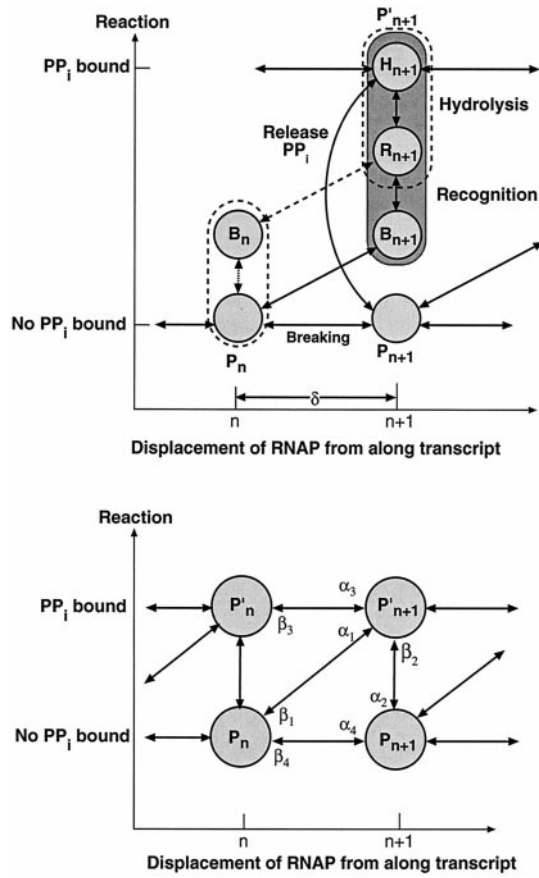
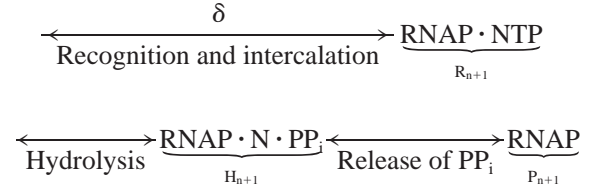
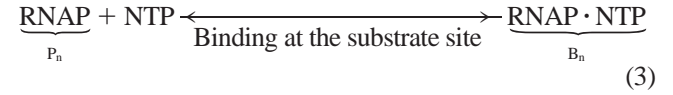


FIGURE 3 (a) The chemical kinetic states for the model shown in Fig. 2. The horizontal axis marks distance in units of base pairs (1 bp = 0.34 nm), and the vertical axis is the reaction coordinate. All transitions with a horizontal component depend on the load force,  $f$ . With the RNA strand in polymerization state  $P_n$ , the catalytic site expands by a distance  $\geq \delta = 0.34$  nm to accommodate the binding of a nucleotide:  $P_n \rightarrow B_{n+1}$ . If the nucleotide is complementary to the template, it binds to form the recognition complex,  $R_{n+1}$ , which triggers rapid hydrolysis to state  $H_{n+1}$ . Release of pyrophosphate carries the system to state  $P_{n+1}$ . If  $PP_i$  release is the rate-limiting step, we can combine states  $B_{n+1}$ ,  $R_{n+1}$ , and  $H_{n+1}$  into a composite state,  $P'_{n+1}$ , shown shaded. An alternative pathway is shown by the dashed transitions. Here binding does not require an expansion of the substrate site, but recognition does. In this case, the state  $P_n$  is the composite state shown enclosed by dashes, and  $P_{n+1}$  contains only  $R_{n+1}$  and  $H_{n+1}$ . (b) The Markov chain model for the transition diagram in a, where we assume that the rate-limiting step is  $PP_i$  release. Thus the transition rates between states  $B_{n+1}$ ,  $R_{n+1}$ , and  $H_{n+1}$  are fast, so that these states can be combined into the single shaded state  $P'_{n+1}$ . All transitions with a horizontal component depend on the load force,  $f$ .

In this scheme, NTP binding to the substrate site requires that the site expand by a distance  $\geq \delta = 0.34$  nm to accommodate the incoming nucleotide. Once docked at the substrate site, a recognition step takes place wherein if the incorrect nucleotide has bound, it is quickly released and another binds. If the nucleotide matches the template, hydrogen bonds are formed between the incoming nucleotide and the template strand, which triggers hydrolysis and subsequent release of pyrophosphate.

Alternatively, binding to the substrate site could occur before intercalation:



In this scheme the catalytic site expands by  $\delta$  to permit recognition only after NTP has bound to the substrate site. This is indicated by the dashed transitions in Fig. 3 a. Other kinetic sequences are also possible. However, we shall sidestep these ambiguities by assuming that under no-load conditions, the rate-limiting step for transcript elongation is the release of pyrophosphate from the catalytic site (Erie et al., 1992; Yin et al., 1995). This enables us to collapse RNAP states connected by fast transitions that are load independent into one composite state, shown enclosed in Fig. 3 a (Scheme 2 by solid lines, Scheme 3 by dashed lines).

With this key assumption about the rate-limiting step, we can represent the state transition diagram in Fig. 3 a by the Markov chain shown in Fig. 3 b. In this simplified model of the polymerization kinetics, an RNAP containing a transcript of length  $n$  can exist in two polymerization states:

- $P_n$  (containing a transcript of length  $n$  with no  $PP_i$  bound)
- $P'_n$  (containing a transcript of length  $n$  with  $PP_i$  bound).

We use in the diagram and the subsequent analysis the notation listed in the Glossary above.

The governing equations for the Markov chain in Fig. 3 b are

$$\frac{d\rho_1(n, t)}{dt} = \rho_1(n+1, t)\beta_4 + \rho_2(n+1, t)\beta_1 + \rho_2(n, t)\alpha_2 + \rho_1(n-1, t)\alpha_4 - \rho_1(n, t)(\alpha_4 + \alpha_1 + \beta_2 + \beta_4)$$

$$\frac{d\rho_2(n, t)}{dt} = \rho_2(n+1, t)\beta_3 + \rho_1(n, t)\beta_2 + \rho_1(n-1, t)\alpha_1 + \rho_2(n-1, t)\alpha_3 - \rho_2(n, t)(\alpha_3 + \alpha_2 + \beta_1 + \beta_3) \quad (4)$$

where  $\sum \rho_1(n, t) + \sum \rho_2(n, t) = 1$ . The solution to these equations will provide the force-velocity curve we seek.

## RESULTS

In Appendix C we solve Eq. 4 corresponding to the model in Fig. 3 b to obtain the following expression for the

load-velocity curve:

$$v = \frac{\alpha_1\alpha_2 - \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 - \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 - \beta_4)}{\alpha_1 + \beta_1 + \alpha_2 + \beta_2} \quad (5)$$

Independent information about the motor function is contained in the statistical variance of the motor's motion about its mean velocity. This variance can be characterized by an "effective" diffusion constant,  $D_{\text{eff}}$ , given by (cf. Appendix D)

$$D_{\text{eff}} = \delta^2 \left[ \frac{\alpha_1\alpha_2 + \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 + \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 + \beta_4) - 2(\alpha_3 - \beta_3)(\alpha_4 - \beta_4)}{2(\alpha_1 + \beta_1 + \alpha_2 + \beta_2)} - (\alpha_1\alpha_2 - \beta_1\beta_2 - (\alpha_1 + \beta_2)(\alpha_4 - \beta_4) - (\beta_1 + \alpha_2)(\alpha_3 - \beta_3)) \right. \\ \left. \times \frac{(\alpha_1\alpha_2 - \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 - \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 - \beta_4))}{(\alpha_1 + \beta_1 + \alpha_2 + \beta_2)^3} \right] \quad (6)$$

A plot of many trajectories should show that the variance of the trajectories increases with time at a rate

$$\text{var}(x(t)) = \langle x(t)^2 \rangle - \langle x(t) \rangle^2 = 2D_{\text{eff}}t \quad (7)$$

(In Appendix D we show that  $D_{\text{eff}}$  contains the same information as the "randomness parameter" defined by Schnitzer and Block to investigate the step size of kinesin (Schnitzer and Block, 1995, 1997).)

## DETERMINATION OF RATE PARAMETERS

In Eqs. 5 and 6 the transition rates with subscripts 1, 3, and 4 depend on the load force  $f$ . The constraint that the model obey detailed balance at chemical equilibrium (which is required for consistency with thermodynamics) requires that (Hill, 1977, 1989)

$$\frac{\alpha_i}{\beta_i} = \exp\left(\frac{-\Delta G_i - f\delta}{k_B T}\right), \quad i = 1, 3, 4 \quad (8)$$

where  $\Delta G_i$  is the free energy drop of the transition process  $i$ . There are two approaches to determining the force dependence of the transition rates. If the potential wells holding the nucleotide in position on the substrate and product sites were known, then one could model these rates by the Kramers rate law (Hanggi et al., 1990; Risken, 1989), and detailed balance is ensured. However, to use this approach, one must know the free energy changes in the whole transition process, not just at the two end points. Alternatively, if only the total free energy drop in the transition process is known, one can calculate one of the parameters, say  $\alpha$ , using experimental results, and obtain the other parameter from Eq. 8. In Appendix E we use the principle of detailed balance and the empirical measurements listed in Table 3,

augmented by two physically reasonable assumptions, to compute the transition rates. The results are shown in Table 1.

## RELATIONSHIP OF THE MODEL TO EXPERIMENTS AND PREDICTIONS

Using the transition rates calculated in Appendix E and listed in Table 1, the model makes the following predictions about the mechanical behavior of RNAP:

1. Fig. 4, *a* and *b*, shows the load-velocity curves for various concentrations of NTP and  $\text{PP}_i$ . The model predicts that the load velocity is concave. This experimental result (M. D. Wang, personal communication) is not used in our calculation of the transition rates in Appendix E; i.e., we did not fit our model to generate this concave load-velocity curve. Rather, the model predicts this concave load-velocity curve based on the parameters obtained from other experimental data.

2. Fig. 5, *a* and *b*, predicts how the stall force and the maximum velocity should respond to changes in the solution concentrations of NTP and  $\text{PP}_i$ . When the concentration of  $\text{PP}_i$  is increased from 1  $\mu\text{M}$  to 1  $\text{mM}$ , the stall force is virtually unchanged, whereas the maximum velocity is reduced by half. This is another experimental result (Yin et al., 1995) that we did not use in our calculation of the transition rates, but is a prediction borne out by experiments. It is also important to point out our predictions for situations where experimental results are currently not available. In particular, our model predicts that, as the concentration of NTP decreases, the stall force decreases by roughly the same percentage as the maximum velocity.

3. In principle, variance measurements on RNAP can be performed in a manner similar to that of measurements made on kinesin (Schnitzer and Block, 1995; Svoboda et al., 1994). In the previous section we showed that the variance can be characterized by an effective diffusion constant:  $\text{var}(x(t)) = \langle x(t)^2 \rangle - \langle x(t) \rangle^2 = 2D_{\text{eff}}t$ . Fig. 6 shows how  $D_{\text{eff}}$

**TABLE 1** Transition rates computed as described in Appendix E from the principle of detailed balance and the empirical parameters listed in Table 3

Transition rates	
$\alpha_1 = \alpha_{10}^0 \cdot [\text{NTP}] \cdot \exp\left(\frac{-f\delta}{k_B T}\right)$	$\alpha_{10}^0 = 10^6 \text{ M}^{-1} \cdot \text{s}^{-1}$
$\beta_1 = 0.21 \text{ s}^{-1}$	
$\alpha_2 = 31.4 \text{ s}^{-1}$	
$\beta_2 = \beta_2^0 \cdot [\text{PP}_i]$	$\beta_2^0 = 10^6 \text{ M}^{-1} \cdot \text{s}^{-1}$
$\alpha_3 = \alpha_3^0 \cdot [\text{NMP}]$	$\alpha_3^0 = 0.31 \text{ M}^{-1} \cdot \text{s}^{-1}$
$\beta_3 = \beta_{30} \cdot \exp\left(\frac{f\delta}{k_B T}\right)$	$\beta_{30} = 0.063 \text{ s}^{-1}$
$\alpha_4 = \alpha_4^0 \cdot [\text{NMP}]$	$\alpha_4^0 = 46.6 \text{ M}^{-1} \cdot \text{s}^{-1}$
$\beta_4 = \beta_{40} \cdot \exp\left(\frac{f\delta}{k_B T}\right)$	$\beta_{40} = 9.4 \text{ s}^{-1}$

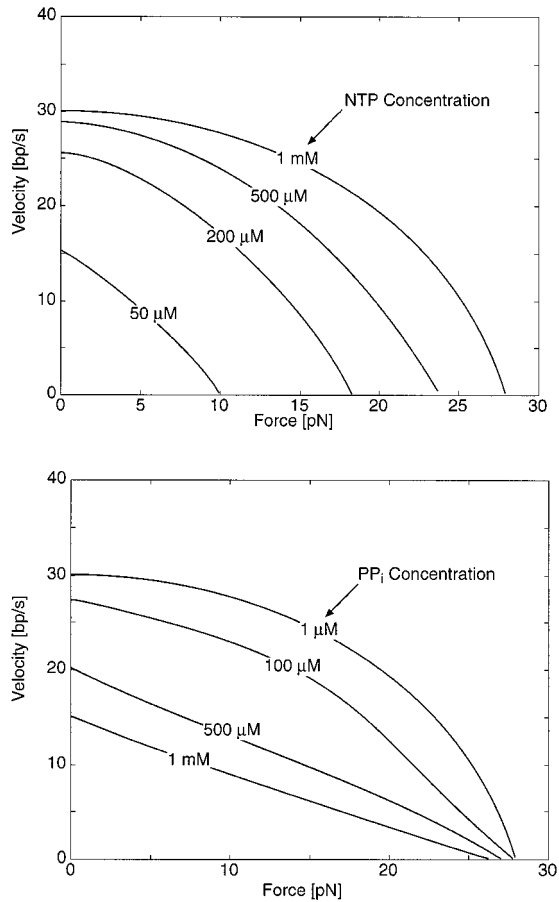


FIGURE 4 (a) Load-velocity curves computed from Eq. 5 for  $1 \mu\text{M}$   $\text{PP}_i$  and different concentrations of NTP. At  $1 \text{ mM}$  NTP and  $1 \mu\text{M}$   $\text{PP}_i$ , the motor initially proceeds almost independently of the resisting force at low loads, then falls sharply to zero at the stall force. As the concentration of NTP decreases, the load-velocity curve becomes less concave. A Java code for the model can be run from within Netscape at [http://teddy.berkeley.edu:1024/Java/ratchet\\_java.html](http://teddy.berkeley.edu:1024/Java/ratchet_java.html). (b) Load-velocity curves computed for  $1 \text{ mM}$  NTP and different concentrations of  $\text{PP}_i$ . Note that as the concentration of  $\text{PP}_i$  is increased from  $1 \mu\text{M}$  to  $1 \text{ mM}$ , the transcription velocity decreases by a factor of 2, whereas the stall force is virtually unchanged.

varies with solution concentrations of NTP and pyrophosphate.

All curves in Figs. 4–6 are computed from Eqs. 5 and 6, using the transition rates given in Table 1.

In Fig. 4 a, at  $1 \text{ mM}$  NTP and  $1 \mu\text{M}$   $\text{PP}_i$ , the motor initially proceeds almost independently of the resisting force at low loads, then falls sharply to zero at the stall force, given by

$$f_s = \frac{k_B T}{\delta} \ln(Q) \approx 28 \text{ pN} \quad (9)$$

where  $Q$  is a steady-state constant given by Eq. E.10 in Appendix E.

For the molecular motors for which a load-velocity curve has been obtained, the relationship is either linear or convex, in contrast to the load-velocity curves for RNAP in Fig. 4 a, which are concave down. At low loads, the RNAP

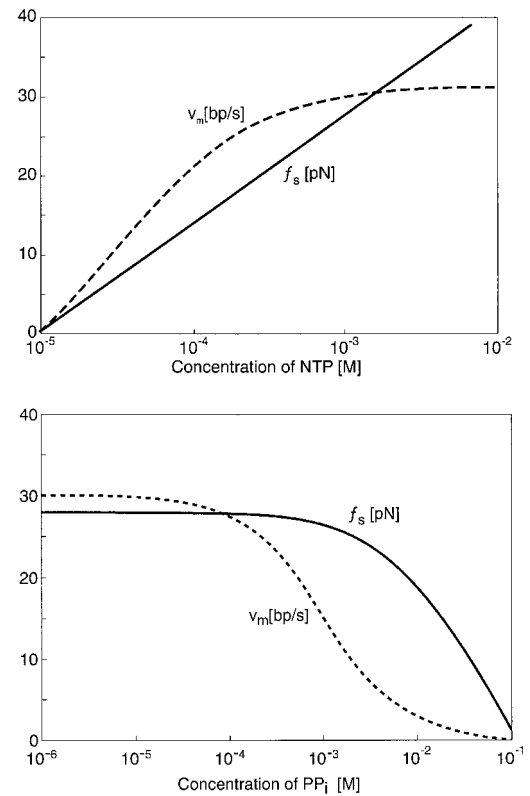


FIGURE 5 (a) The maximum velocity ( $v_m$ ) and the stall force ( $f_s$ ) as functions of NTP concentration at  $1 \mu\text{M}$   $\text{PP}_i$ .  $f_s$  is computed from Eq. E.9. At low NTP concentrations both  $v_m$  and  $f_s$  rise logarithmically with [NTP]; however, at higher concentrations,  $f_s$  continues to rise, whereas  $v_m$  levels off and is nearly constant. (b) The maximum velocity ( $v_m$ ) and the stall force ( $f_s$ ) as functions of  $\text{PP}_i$  concentration at  $1 \text{ mM}$  NTP. At low pyrophosphate concentrations,  $v_m$  and  $f_s$  are practically constant; at moderate pyrophosphate concentrations,  $f_s$  remains constant, but  $v_m$  starts to fall off, first logarithmically, then at a slower pace; at higher concentrations,  $f_s$  decreases logarithmically and converges to zero along with  $v_m$ .

transcription velocity does not decrease significantly as the load force increases.

In general, such a concave shape will arise when there is a rate-limiting chemical step that is not affected by the load force. This can be seen by considering a process consisting of cycles of two sequential steps. Suppose that the rate of the first step decreases exponentially with the load, but the rate of the second step is load independent and is much smaller than that of the first step at zero load. At low loads, the second step is the rate-limiting step; that is, the rate of the process cycle is roughly the same as the second step, no matter how fast the first step is. When the load is increased sufficiently, the rate of the first step eventually decreases to values comparable to that of the second step, whereupon the rate of the process cycle drops sharply to zero. This leads to a concave load velocity curve.

For RNAP during normal progression, the rate-limiting chemical event is pyrophosphate release, which we have assumed is independent of the load. In a typical transcription cycle, polymerization is followed by the release of

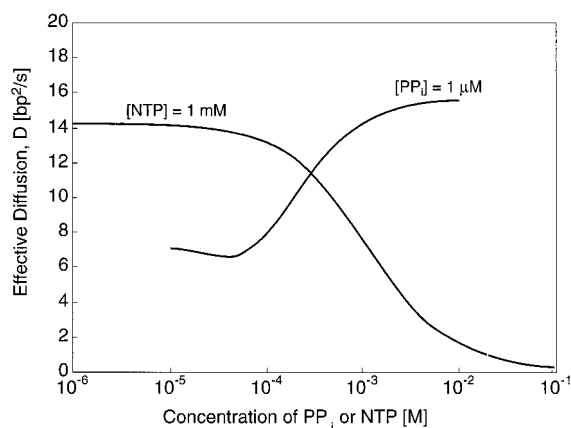


FIGURE 6 The effective diffusion constant ( $D_{\text{eff}}$ ) computed from Eq. 6 as a function of NTP concentration at  $1 \mu\text{M}$   $\text{PP}_i$ , and  $D_{\text{eff}}$  as a function of  $\text{PP}_i$  concentration at  $1 \text{ mM}$  NTP.

pyrophosphate. At low loads, the polymerization step is much faster than the release of  $\text{PP}_i$ , and so the transcription rate is approximately the rate of  $\text{PP}_i$  release. At high loads the polymerization rate eventually falls below the rate of  $\text{PP}_i$  release, whereupon the transcription rate is roughly given by the polymerization rate, which falls off significantly with load.

The stall force  $f_s \approx 28 \text{ pN}$  is well below the thermodynamic limit of  $f_s^{\text{TD}} = \Delta G/\delta \approx 145 \text{ pN}$ , where  $\Delta G \approx 12k_B T = 50 \text{ pN}\cdot\text{nm}$  is the mean free energy of NTP hydrolysis, and  $\delta = 0.34 \text{ nm}$  is the step size (Yin et al., 1995; M. D. Wang, personal communication). Because the thermodynamic maximum stall force  $f_s^{\text{TD}} \approx 1/(\text{step size})$ , the large stall force of RNAP may be attributed to the small step size of the RNAP. The stall force of RNAP is 5–6 times larger than that of myosin or kinesin. However, if we measure the efficiency of energy conversion near stall of each forward step by the ratio  $f_s\delta/\Delta G$ , then we find that RNAP is significantly less efficient than myosin or kinesin. These three molecular motors are compared in Table 2. Note that at stall the RNAP is still hydrolyzing NTP at a steady rate, so energy consumption continues, although no work is being performed. Thus the hydrolysis of NTP and the transcript elongation are not tightly coupled.

### Dependence on pyrophosphate and NTP concentrations

The two quantities that are easiest to manipulate experimentally are the concentrations of pyrophosphate and nucleo-

tide. Fig. 5 shows how the stall force ( $f_s$ ) and the maximum velocity ( $v_m$ ) depend on these variables:

At low NTP concentrations, both  $v_m$  and  $f_s$  rise logarithmically with [NTP]; however, at higher concentrations,  $f_s$  continues to rise, whereas  $v_m$  levels off and is nearly constant.

At low pyrophosphate concentrations,  $v_m$  and  $f_s$  are practically constant. At moderate pyrophosphate concentrations,  $f_s$  remains constant, but  $v_m$  starts to fall off, first logarithmically, then at a slower pace; at higher concentrations,  $f_s$  decreases logarithmically and converges to zero along with  $v_m$ .

### Sequence dependence

The model developed so far treats DNA as a homopolymer with but one nucleotide type. There are several ways to generalize the model to include sequence-specific effects. The most important mechanical parameter is the strength of the bonds holding the terminal nucleotide onto the end of the transcript. This is embodied in the “horizontal” dissociation rate constants,  $\beta_i$ , in Fig. 3. If all other factors are equal, the  $\beta$ 's can take on one of two values: A-U and A-T pairs are joined by two hydrogen bonds (denote by  $\beta''$ ), and G-C pairs are held together by a triplet of hydrogen bonds (denote by  $\beta'''$ ). RNAP processing along a homopolymeric DNA consisting of all G's will have a higher stall force than one consisting of all A's, because the breaking rate,  $\beta$ 's, will be smaller for the former than for the latter (cf. Appendix E). The stall force of a DNA strand consisting of a random sequence of bases will have a mixture of two values of  $\beta = (\beta'', \beta''')$ . The average stall force will lie between two extremes:  $f_s(\beta'') \leq \langle f_s \rangle \leq f_s(\beta''')$ .

For a given DNA sequence, the Markov chain in Fig. 3 *b* can be simulated by replicating each Markov unit with a  $\beta$  ( $\beta''$  or  $\beta'''$ ) corresponding to the bond type at that location. In this way, a distribution of stall loads will be computed for many replications of the numerical “experiment.” No data are currently available for sequence-dependent stall loads; however, the model makes definite predictions of how such measurements should go.

### Backsliding

We have viewed RNAP as a processive “sliding clamp” with the 3' terminus of the RNA transcript always aligned with the catalytic site (Landick, 1997). However, there is evidence that the RNA transcript can slip out of the catalytic site, allowing the RNAP transcription bubble to backslide along the RNA and DNA, a phenomenon that is probably

TABLE 2 Comparison of RNAP with kinesin and myosin

	Step size	Stall force	$\Delta G_{\text{HYDROL}}$	Efficiency	Reference
RNAP	0.34 nm	28 pN	$12 k_B T$	0.2	M. D. Wang, personal communication
Kinesin	8 nm	6 pN	$20 k_B T$	0.6	Svoboda and Block (1994)
Myosin	11 nm	3–5 pN	$20 k_B T$	0.6	Finer et al. (1994)

sequence dependent (Landick, 1997). Moreover, in the laser trap force measurements, RNAP frequently paused for variable amounts of time before resuming transcription and force generation (M. D. Wang, personal communication). This also may be due to slipping of the leading RNA nucleotide out of the catalytic site. Such effects can be incorporated into the model, as shown in Fig. 7, by introducing a parallel sequence of states that allow backsliding of the RNAP along the RNA and DNA, which is probably force dependent.

### The appearance of “inchworming” in DNA footprinting

According to the model as formulated, the motion of the RNAP proceeds “smoothly,” one step at a time, which is in accord with some recent observations (Nudler et al., 1997). However, there is also evidence from DNA footprinting studies that RNAP “inchworms” along the DNA in a saltatory motion (Chamberlin, 1994; Krummel and Chamberlin, 1992). Such motion can be accommodated into the model by allowing the posterior subunit of the RNAP to be elastically tethered to the front subunit, as shown schematically

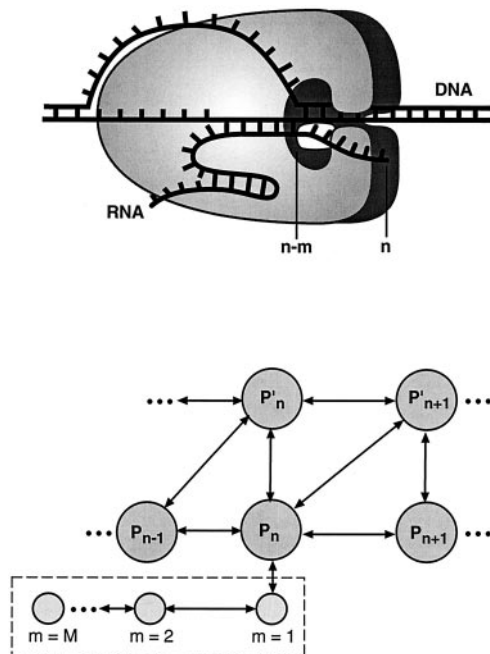


FIGURE 7 (a) Schematic diagram of RNAP backsliding. The RNA transcript slips out of the catalytic site, allowing the RNAP transcription bubble to backslide along the RNA and DNA. The catalytic site at location  $n$  slips one or more nucleotides back along the transcript, so that the 3' terminus of the transcript is still at position  $n$ , but the catalytic site is now at position  $n - m$ . (b) Markov chain model for simulating RNAP backsliding. The top portion of the diagram is the same as in Fig. 3 b. The RNAP starts backsliding at position  $n$ . The dashed box represents states where the 3' terminus of the RNA transcript is not aligned with the catalytic site. These states are numbered by  $m = 1, 2, \dots, M$ , where  $M$  is the number of base pairs that have slipped out of the catalytic site.

in Fig. 2. This can produce an apparent variation in the RNAP footprint by two different mechanisms.

First, suppose that the attachment of the transcript to the posterior subunit can be modeled as a series of potential wells representing the attachments (e.g., hydrogen bonds) of the transcript to the RNAP. As the front end moves steadily forward, the spring is stretched and the force on the rear increases until the threshold for breaking the bonds is reached. Thereafter, the rear end will process at the same rate as the front end, but the overall dimension of the RNAP will be dilated. When the progress of the front end is terminated during the footprinting assay, the RNAP will relax back to its equilibrium length, and the assay will give the impression of a variable-length protected region of the DNA. This length will be sequence dependent, because the strength of the attachment of the transcript to the RNAP is sequence dependent.

Alternatively, if the binding of the transcript to the rear end has the character of a “sliding friction,” then the rear end will be stationary until a threshold force is exceeded, whereupon it will lurch forward to its equilibrium position. The front end will then extend the spring once again. Thus inchworming of the rear end of the RNAP will accompany the smooth motion of the front end.

### DISCUSSION

As RNA polymerase carries out transcription under cellular conditions, it creates downstream supercoiling that generates an opposing load force of  $\sim 6$  pN (Yin et al., 1995). However, laser trap experiments have shown that it can move against a load several times greater than the maximum force developed by kinesin (M. D. Wang, personal communication; Svoboda and Block, 1994). Moreover, the load-velocity curve appears to differ qualitatively from that measured for kinesin and myosin. Rather than falling off almost linearly with load, RNAP moves at nearly constant velocity until loads above  $\sim 20$  pN are applied, whereupon it falls off rapidly to a stall force of 25–30 pN.

Because the details of nucleotide binding, hydrolysis, and polymerization are not known exactly, we have formulated a kinetic model based on energetics that allows us to understand certain qualitative and quantitative aspects of RNAP's mechanochemistry. Each step of the transcription process consists of several kinetic steps whose transition rates are calculated based on experimentally determined free energy differences. The energy driving RNAP motion derives ultimately from the hydrolysis and subsequent binding of the nucleotide triphosphates it uses to build the RNA transcript. However, it is not clear how this free energy is transduced into so large a processive force. This transduction process cannot be too efficient, for the measured stall force is much less than the thermodynamically maximum force obtained by dividing the free energy of hydrolysis by the length of a single base:  $\sim 12k_B T$  (at  $1 \mu\text{M}$  PP<sub>i</sub> and  $1 \text{ mM}$  NTP)/ $0.34 \text{ nm} \approx 145 \text{ pN}$  (Yin et al., 1995).



The model presented here provides an explanation for this energy transduction; it is essentially an extension of the Brownian ratchet polymerization models developed earlier (Mogilner and Oster, 1996; Peskin et al., 1993). A related idea for RNAP procession was proposed by Yager and von Hippel (1987); however, we are not aware of any quantitative model for RNAP force generation. The model demonstrates that a Brownian ratchet mechanism with a step size of one nucleotide is sufficient to account for the sizable stall force measured in the laser trap experiments. No conformational changes need be invoked, although the model does not rule them out. Indeed, in a previous model for kinesin force generation, it was shown that a combination of biased diffusion and a conformation change driven by binding free energy was necessary to reproduce the observed mechanical measurements (Peskin and Oster, 1995).

In addition to accounting for the large observed stall force, the model accounts for the atypical concave shape of the load-velocity curve. This shape arises because of the multiple chemical steps involved in the ratchet mechanism. In particular, this concave shape is essentially due to the slow (relative to the mechanical motions) rate of pyrophosphate release, which is the rate-limiting step in transcription. Because the time scale for mechanical relaxation to equilibrium is so much faster than the rate of the fastest reactions involved, the progression of RNAP can be treated as a sequence of mechanical equilibrium states and modeled as a Markov chain with thermally excited transition rates. A similar viewpoint was taken by Erie et al. in discussing the kinetics of transcription (Erie et al., 1992).

We have restricted ourselves to the simplest case of two kinetic steps: polymerization and the rate-limiting step of pyrophosphate release. However, there is no difficulty in generalizing the model to include other kinetic schemes. It is also relatively straightforward to include sequence-dependent rates once such data become available.

Experimentally determining the load-velocity curve for RNAP has proved difficult because of the occurrence of confounding effects. For example, RNAP occasionally slides backward, resulting in the leading nucleotide temporarily slipping out of the catalytic region. This backsliding is much slower than a normal transcription step, and so, when it occurs, backsliding becomes the rate-limiting step in the overall transcription. Between the backsliding events, the RNAP moves with roughly constant velocity. Thus the velocity of RNAP is constant when it is "on the track," but drops to zero when it slides "off the track." When it slides back on the track, the velocity jumps up again to its normal rate. Thus the "net" velocity of transcription when the RNAP is on the track is given by the envelope of the RNAP velocity (M. D. Wang, personal communication). In each step of transcription between backslidings, the release of pyrophosphate is the rate-limiting step. Although our model was developed for the normal transcription where the RNAP is on the track, it can be extended to take the backslidings into consideration (see Fig. 7 and the accompanying discussion).

We have avoided explicit treatment of the controversial issue of inchworm motion by RNAP (Chamberlin, 1994; Nudler et al., 1997), although we have indicated how this feature can be included by modeling RNAP as two elastically joined subunits and taking into account the binding of the transcript to the posterior section. This addition also accounts for pausing by upstream RNA hairpin formation.

Finally, the model makes definite predictions about how the stall force depends on pyrophosphate and nucleotide triphosphate concentrations, how the various kinetic steps reflect on the load-velocity behavior, and how measuring the statistics of progression can provide information about the kinetics of transcription. Thus the model can serve as a basis for a unified view of the kinetics, thermodynamics, and mechanics of transcription by RNAP.

## APPENDICES

### A. Relaxation to mechanical equilibrium

Here we estimate the relaxation time of the system to the equilibrium state. We show that the time scale for the system to relax to mechanical equilibrium after a successful polymerization is much smaller than the time scale of polymerization. Therefore, the system is always in thermodynamic equilibrium with respect to the polymerization process.

The largest fluctuating element in the system is the bead in the laser trap. So the fluctuation of the bead is much smaller than the fluctuations of other elements in the system, i.e.,  $D_{\text{bead}} \ll D_{\text{RNAP}}$ . Thus the relaxation time we shall compute based on the fluctuation of the bead is definitely an overestimate of the actual relaxation time of the RNAP.

The bead used in the laser trap experiments was  $0.52 \mu\text{m}$  in diameter (Yin et al., 1995). Using Stokes' law and the Einstein relationship, its diffusion coefficient in water is

$$D = \frac{k_B T}{6\pi\eta r} \approx 8.4 \times 10^5 \text{ nm}^2 \cdot \text{s}^{-1} \quad (\text{A.1})$$

Let  $x$  denote the distance between the transcript tip and the RNAP barrier (labeled F in Fig. 2). Let  $\text{Prob}(x > \delta; t)$  be the probability that  $x > \delta$  at time  $t$ . Suppose that at  $t = 0$ , the distance between the RNA-DNA tip and the RNAP barrier is zero. Because it takes an infinite amount of time to relax to the exact equilibrium state, we define the mechanical relaxation time  $t^*$  as

$$\text{Prob}(x > \delta; t^*) \equiv 0.9 \times \text{Prob}(x > \delta; \infty) \quad (\text{A.2})$$

Note that the relaxation time  $t^*$  is defined in terms of the probability of having fluctuations greater than  $\delta$ , which is the probability affecting polymerization.

Let  $\phi(x, t)$  be the probability density that the distance between the RNA tip and the RNAP barrier is  $x$  at time  $t$ . The governing equation for  $\phi(x, t)$  is

$$\frac{\partial \phi}{\partial t} = D \frac{\partial^2 \phi}{\partial x^2} + D \frac{f}{k_B T} \frac{\partial \phi}{\partial x} \quad (\text{A.3})$$

where  $D$  is the diffusion coefficient of the RNA tip relative to the RNAP barrier, and  $f$  is the load force. Introducing nondimensional length and time scales,

$$y = \frac{x}{\delta}; \quad \tau = t \left( \frac{D}{\delta^2} \right), \quad (\text{A.4})$$

the nondimensional version of Eq. A.3 reads

$$\frac{\partial \phi}{\partial \tau} = \frac{\partial^2 \phi}{\partial y^2} + \omega \frac{\partial \phi}{\partial y}, \quad \omega = \frac{f \delta}{k_B T} \quad (\text{A.5})$$

with boundary conditions  $(\partial \phi / \partial y)|_{y=0} = 0$ , and  $\int_0^\infty \phi dy = 1$ .  
The equilibrium solution of Eq. A.5 is given by

$$\phi(y, \infty) = \frac{1}{\omega} e^{-\omega y} \quad (\text{A.6})$$

Let  $\tau^*$  be the nondimensional time to relax to within 10% of equilibrium:

$$\int_1^\infty \phi(y, \tau^*) dy = 0.9 \times \int_1^\infty \phi(y, \infty) dy = 0.9 \times e^{-\omega} \quad (\text{A.7})$$

According to the definition, the (dimensional) relaxation time  $t^*$  of the system can be expressed in terms of  $\tau^*$  as

$$t^*[\text{s}] = \tau^* \frac{\delta^2}{D} = \tau^* \cdot (1.4 \times 10^{-7}) \quad (\text{A.8})$$

To see the behavior of  $t^*(f)$  for  $f \geq 0$ , we solve Eq. A.5 for different values of  $\omega$ . At each time step, the integral of  $\phi(y, \tau^n)$  over  $(1, \infty)$  is compared with its equilibrium value. When the integral is within 10% of its equilibrium value,  $\tau^* = \tau^n$  is recorded and  $t^*$  is calculated using Eq. A.8.

In Fig. 8 we plot the relaxation time  $t^*$  as a function of the load force. Two features of the plot are worth noting:

The relaxation time is on the order of  $10^{-6}$  s. This justifies the assumption that the system relaxation time is small in comparison with the time scale of polymerization.

The relaxation time is a decreasing function of the load force. This may seem counterintuitive at first; however, it becomes clear when we realize that the load force affects the relaxation time in two opposite ways. On the one hand, the load force makes it hard for the transcript tip to fluctuate away from the RNAP barrier  $F$ ; this slows down the relaxation process. On the other hand, the load force moves the equilibrium state closer to the initial state; that is, the equilibrium state corresponding to a large load force is not far from the initial state. This reduces the ‘‘amount of relaxation’’ the system has to do to reach equilibrium. From Fig. 8 it is clear that the second factor dominates the first one, and so the relaxation time decreases as the load force increases. It is worth noticing that the load force reduces the

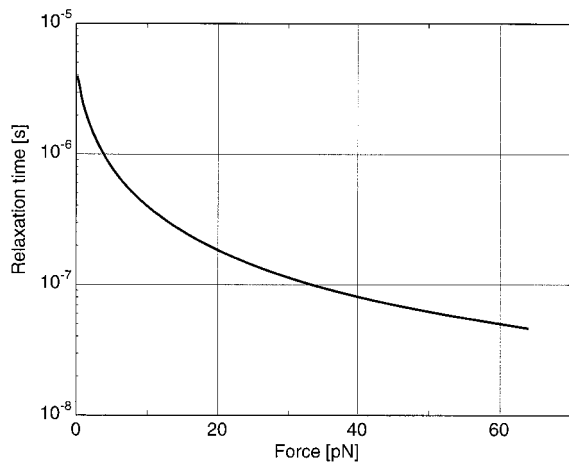


FIGURE 8 The relaxation time of the system to mechanical equilibrium as a function of the load force. The relaxation time is on the order  $10^{-6}$  s, much smaller than the time scale of polymerization. So the system is always in equilibrium, as far as the polymerization process is concerned.

relaxation time, not by increasing the speed at which the system approaches equilibrium, but by pulling the final equilibrium distribution closer to the initial distribution.

## B. All elasticities in the system can be modeled by an effective constant force

### Two springs in series

Here we study the equilibrium distribution of displacements of two springs in series. We show that, at equilibrium, two springs in series can be replaced by a single effective spring.

Consider two objects,  $A_1$  and  $A_2$ , connected in series through a third object,  $A_3$ , by two springs with elastic constants  $k_1$  and  $k_2$ . Let  $x_i$  denote the position of  $A_i$  ( $i = 1, 2, 3$ ). The equilibrium distribution of  $(x_1, x_2, x_3)$  is given by

$$\rho(x_1, x_2, x_3) \propto \exp\left(\frac{-k_1(x_1 - x_3)^2 - k_2(x_2 - x_3)^2}{2k_B T}\right) \quad (\text{B.1})$$

We obtain the joint equilibrium distribution of  $(x_1, x_2)$  by integrating Eq. B.1 with respect to  $x_3$ :

$$\rho(x_1, x_2) \propto \exp\left(\frac{-k_{12}(x_1 - x_2)^2}{2k_B T}\right) \quad (\text{B.2})$$

Equation B.2 indicates that the equilibrium distribution of  $(x_1, x_2)$  is the same as if the objects  $A_1$  and  $A_2$  are connected by an effective spring with elastic constant  $k_{12}$  given by

$$k_{12} = \frac{k_1 k_2}{k_1 + k_2} \leq \min(k_1, k_2) \quad (\text{B.3})$$

### The effective elasticity of the RNAP system

Fig. 9 shows the mechanical arrangement in the load-velocity experiment. The bead is held in the laser trap. This is equivalent to a linear spring with elastic constant  $k_1$ , which is fixed to the substratum. The tip of the RNA-DNA hybrid is connected through the DNA strand to the bead; spring  $k_2$  represents the elasticity in the DNA strand. The barrier  $F$  on RNAP is always downstream from the RNA tip and is mechanically connected to the substratum through RNAP. Spring  $k_3$  represents the elasticity in the connection between the barrier and the substratum.

Because springs  $k_1$  and  $k_2$  are in series, we can combine them into an effective spring,  $k_{12}$ , given by Eq. B.3. It follows from B.3 that  $k_{12} < k_1$ ,

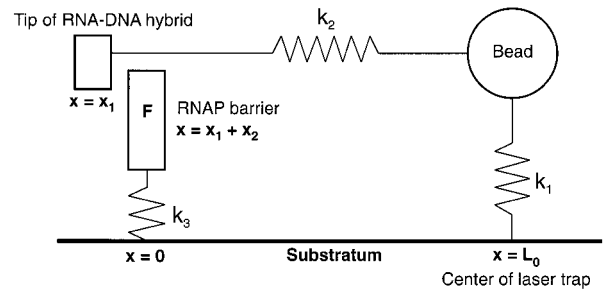


FIGURE 9 The mechanical arrangement of the laser trap experiments. The bead is held in the laser trap (spring  $k_1$ ), which is fixed to the substratum. The tip of the DNA-RNA hybrid is linked through the DNA strand to the bead; spring  $k_2$  represents the elasticity in the DNA strand. The barrier on RNAP is connected to the substrate through RNAP. Spring  $k_3$  represents the elasticity of the RNAP protein connecting the barrier and the substrate. The system is constrained by the condition  $x_2 > 0$ .

i.e., the effective spring  $k_{12}$  is weaker than the spring constant of the laser trap. Let us introduce the coordinates:

$x = 0$  is the point where the RNAP is tethered to the substrate.

The downstream direction is defined as the positive direction.

$x = L_0$  is the center of the laser trap.

$x = x_1$  is the tip of the RNA-DNA hybrid.

$x = x_1 + x_2$  is the position of the barrier on RNAP.

Note that the distance from the RNA tip to the barrier is  $x_2$ ; therefore, the system is constrained by the condition  $x_2 > 0$ . Let  $L_{12}$  be the rest length of spring  $k_{12}$ . Without loss of generality, we can assume that the rest length of spring  $k_3$  is zero. The elastic energy of the system is given by

$$E(x_1, x_2) = \frac{k_{12}}{2} (L_0 - x_1 - L_{12})^2 + \frac{k_3}{2} (x_1 + x_2)^2 \quad (\text{B.4})$$

The equilibrium distribution of  $(x_1, x_2)$  is then

$$\rho(x_1, x_2) \propto \exp\left(\frac{-k_{12}(L_0 - x_1 - L_{12})^2 - k_3(x_1 + x_2)^2}{2k_B T}\right) \quad (\text{B.5})$$

### The effective constant force

In the force-velocity experiments, the spring constant  $k_1$  of the laser trap is between 0.03 and 0.2 pN/nm (Yin et al., 1995), which corresponds to

$$\frac{k_1 \delta^2}{2k_B T} = 4.2 \times 10^{-4} \text{ to } 2.8 \times 10^{-3} \quad (\text{B.6})$$

Because the effective spring  $k_{12}$  is even weaker than spring  $k_1$ , we can safely assume that spring  $k_{12}$  is weak; more specifically, we assume that

$$\frac{k_{12} \delta^2}{2k_B T} \ll 1 \quad (\text{B.7})$$

We now show that, under assumption B.7, the probability that  $x_2 > \delta$  is given approximately by

$$P(x_2 > \delta) = \exp\left(\frac{-\langle f \rangle \delta}{k_B T}\right) \quad (\text{B.8})$$

where  $\langle f \rangle$  is the (Boltzmann) average force exerted on the tip of the RNA-DNA hybrid.

To proceed, we first examine the equilibrium distribution of  $x_1$ :

$$\begin{aligned} \rho(x_1) &= \int_0^\infty \rho(x_1, x_2) dx_2 \\ &\propto \exp\left(\frac{-k_{12}(L_0 - x_1 - L_{12})^2}{2k_B T}\right) \cdot \text{erfc}\left(\sqrt{\frac{k_3}{2k_B T}} x_1\right) \end{aligned} \quad (\text{B.9})$$

where  $\text{erfc}(r)$  is the complementary error function defined by

$$\text{erfc}(r) \equiv \frac{2}{\sqrt{\pi}} \int_r^\infty \exp(-s^2) ds \quad (\text{B.10})$$

From Eq. B.9 it can be shown that

$$\text{var}(x_1) < \frac{k_B T}{k_{12}} \quad (\text{B.11})$$

This result is not surprising, for if the RNA tip is only restricted by spring  $k_{12}$  (i.e., the RNAP barrier is not present), the variance of  $x_1$  is exactly

$k_B T/k_{12}$ . The presence of the RNAP barrier introduces additional restrictions on  $x_1$ , thus reducing its variance.

Next we calculate the probability that  $x_2 > \delta$ :

$$P(x_2 > \delta) = \frac{\int_\delta^\infty dx_2 \int_{-\infty}^\infty dx_1 \rho(x_1, x_2)}{\int_0^\infty dx_2 \int_{-\infty}^\infty dx_1 \rho(x_1, x_2)} \quad (\text{B.12})$$

Let  $\langle x_1 \rangle$  denote the mean of  $x_1$ . Using the substitutions  $y_1 = x_1 + \delta$  and  $y_2 = x_2 - \delta$  in the numerator, we can write Eq. B.12 as

$$\begin{aligned} P(x_2 > \delta) &= \frac{\int_0^\infty dy_2 \int_{-\infty}^\infty dy_1 \exp\left(\frac{-k_{12}(L_0 - y_1 + \delta - L_{12})^2 - k_3(y_1 + y_2)^2}{2k_B T}\right)}{\int_0^\infty dx_2 \int_{-\infty}^\infty dx_1 \exp\left(\frac{-k_{12}(L_0 - x_1 - L_{12})^2 - k_3(x_1 + x_2)^2}{2k_B T}\right)} \\ &= \exp\left(-\frac{k_{12} \delta^2}{2k_B T} - \frac{k_{12}(L_0 - \langle x_1 \rangle - L_{12}) \delta}{k_B T}\right) \\ &\quad \cdot \int_0^\infty dy_2 \int_{-\infty}^\infty dy_1 \rho(y_1, y_2) \exp\left(\frac{k_{12} \delta}{k_B T} (y_1 - \langle x_1 \rangle)\right) \end{aligned} \quad (\text{B.13})$$

Expanding the integrand in Eq. B.13 in a Taylor series, we have

$$\begin{aligned} \rho(y_1, y_2) \cdot \exp\left(\frac{k_{12} \delta}{k_B T} (y_1 - \langle x_1 \rangle)\right) &= \rho(y_1, y_2) \\ &\quad \cdot \left(1 + \frac{k_{12} \delta}{k_B T} (y_1 - \langle x_1 \rangle) + \frac{1}{2} \left(\frac{k_{12} \delta}{k_B T}\right)^2 (y_1 - \langle x_1 \rangle)^2 + \dots\right) \end{aligned} \quad (\text{B.14})$$

Substituting Eq. B.14 into Eq. B.13 yields

$$\begin{aligned} P(x_2 > \delta) &= \exp\left(-\frac{k_{12} \delta^2}{2k_B T} - \frac{k_{12}(L_0 - \langle x_1 \rangle - L_{12}) \delta}{k_B T}\right) \\ &\quad \cdot \left(1 + \frac{k_{12} \delta}{k_B T} (\langle x_1 \rangle - \langle x_1 \rangle) + \frac{1}{2} \left(\frac{k_{12} \delta}{k_B T}\right)^2 \text{var}(x_1) + \dots\right) \end{aligned} \quad (\text{B.15})$$

Using inequality B.11 and assumption B.7, we obtain

$$P(x_2 > \delta) = \exp\left(-\frac{\langle f \rangle \delta}{k_B T}\right) \cdot \left(1 + \mathcal{O}\left(\frac{k_{12} \delta^2}{2k_B T}\right)\right) \quad (\text{B.16})$$

where

$$\langle f \rangle = k_{12}(L_0 - \langle x_1 \rangle - L_{12}) \quad (\text{B.17})$$

is the average force exerted on the tip of the RNA-DNA hybrid.

## C. The load-velocity curve

In this appendix we derive the expression (Eq. 5) for the average velocity of the RNA tip. The mean position of the RNA tip (measured in base pairs from the origin at the beginning of the RNA chain) is given by

$$\langle x(t) \rangle = \sum_n n (\rho_1(n, t) + \rho_2(n, t)) \quad (\text{C.1})$$

From Eq. 4 we have

$$\begin{aligned} v &= \frac{d\langle x(t) \rangle}{dt} = \sum_n n \left( \frac{d\rho_1(n, t)}{dt} + \frac{d\rho_2(n, t)}{dt} \right) \\ &= (\alpha_1 + \alpha_4 - \beta_4) \sum_n \rho_1(n, t) - (\beta_1 + \beta_3 - \alpha_3) \sum_n \rho_2(n, t) \\ &\equiv (\alpha_1 + \alpha_4 - \beta_4) \rho_1(t) - (\beta_1 + \beta_3 - \alpha_3) \rho_2(t) \end{aligned} \quad (\text{C.2})$$

where

$$\rho_1(t) = \sum_n \rho_1(n, t), \quad \rho_2(t) = \sum_n \rho_2(n, t)$$

are, respectively, the probability that no PP<sub>i</sub> is bound on the RNAP, and the probability that a PP<sub>i</sub> is bound on the RNAP. The probabilities  $\rho_1(t)$  and  $\rho_2(t)$  satisfy

$$\rho_1(t) + \rho_2(t) = 1 \quad (\text{C.3})$$

Taking the sum of both sides of Eq. 4, we obtain

$$\frac{d\rho_1(t)}{dt} = \rho_2(t)(\alpha_2 + \beta_1) - \rho_1(t)(\alpha_1 + \beta_2) \quad (\text{C.4})$$

$$\frac{d\rho_2(t)}{dt} = \rho_1(t)(\alpha_1 + \beta_2) - \rho_2(t)(\alpha_2 + \beta_1) \quad (\text{C.5})$$

The solution to Eqs. C.4 and C.5 converges to the steady-state solution very fast: the relaxation time to the steady-state solution is on the order of  $1/(\alpha_1 + \beta_1 + \alpha_2 + \beta_2)$ . Whatever the initial values of  $\rho_1(t)$  and  $\rho_2(t)$ , after a very short time,  $\rho_1(t)$  and  $\rho_2(t)$  are essentially equal to the steady-state solution. Because the average velocity is a steady-state quantity, we use the steady-state values of  $\rho_1(t)$  and  $\rho_2(t)$  in the calculation of the average velocity. The steady-state solution satisfies

$$\rho_2(\alpha_2 + \beta_1) - \rho_1(\alpha_1 + \beta_2) = 0 \quad (\text{C.6})$$

Solving for  $\rho_1$  and  $\rho_2$ , we obtain

$$\begin{aligned} \rho_1 &= \frac{\alpha_2 + \beta_1}{\alpha_1 + \beta_1 + \alpha_2 + \beta_2} \\ \rho_2 &= \frac{\alpha_1 + \beta_2}{\alpha_1 + \beta_1 + \alpha_2 + \beta_2} \end{aligned} \quad (\text{C.7})$$

The average velocity,  $v$ , in bp/s, is given by

$$\begin{aligned} v &= (\alpha_1 + \alpha_4 - \beta_4) \rho_1(t) - (\beta_1 + \beta_3 - \alpha_3) \rho_2(t) \\ &= \frac{\alpha_1 \alpha_2 - \beta_1 \beta_2 + (\alpha_1 + \beta_2)(\alpha_3 - \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 - \beta_4)}{\alpha_1 + \beta_1 + \alpha_2 + \beta_2} \end{aligned} \quad (\text{C.8})$$

The transition rates in Eq. C.8 are given in Table 1.

## D. Variance analysis of the trajectories

As polymerization proceeds, the dispersion in the position of the RNA tip increases because of the stochastic nature of the motion reaction processes. This creates an effective diffusion in the spatial direction, which we can characterize by an effective diffusion coefficient,  $D$ . We can derive an

expression for  $D$  by examining the evolution of a solution of the form

$$\begin{bmatrix} \rho_1(n, t) \\ \rho_2(n, t) \end{bmatrix} = e^{ik(n\delta)} \begin{bmatrix} c_1(t) \\ c_2(t) \end{bmatrix} \quad (\text{D.1})$$

Substituting Eq. D.1 into the Markov chain equations (Eq. 4), we obtain the evolution equation for  $c_1(t)$ ,  $c_2(t)$ :

$$\frac{d}{dt} \begin{bmatrix} c_1(t) \\ c_2(t) \end{bmatrix} = \mathbf{A} \begin{bmatrix} c_1(t) \\ c_2(t) \end{bmatrix} \quad (\text{D.2})$$

where

$$\mathbf{A} = \begin{bmatrix} (e^{ik\delta} - 1)\beta_4 + (e^{-ik\delta} - 1)\alpha_4 - (\alpha_1 + \beta_2), & (e^{ik\delta} - 1)\beta_1 + (\beta_1 + \alpha_2) \\ (e^{-ik\delta} - 1)\alpha_1 + (\alpha_1 + \beta_2), & (e^{ik\delta} - 1)\beta_3 + (e^{-ik\delta} - 1)\alpha_3 - (\beta_1 + \alpha_2) \end{bmatrix} \quad (\text{D.3})$$

The eigenvalues of the matrix  $\mathbf{A}$  are the roots of the characteristic polynomial:

$$\begin{aligned} 0 &= \det(\lambda \mathbf{I} - \mathbf{A}) \\ &= \lambda^2 + \lambda(\alpha_1 + \beta_1 + \alpha_2 + \beta_2 + ik\delta(\alpha_3 + \alpha_4 - \beta_3 - \beta_4) \\ &\quad + O((k\delta)^2)) \\ &\quad + ik\delta(\alpha_1\alpha_2 - \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 - \beta_3) + (\beta_1 + \alpha_2) \\ &\quad \cdot (\alpha_4 - \beta_4)) \\ &\quad - (k\delta)^2 \left( (\alpha_3 - \beta_3)(\alpha_4 - \beta_4) \right. \\ &\quad \left. - \frac{\alpha_1\alpha_2 + \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 + \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 + \beta_4)}{2} \right) + O((k\delta)^3) \end{aligned} \quad (\text{D.4})$$

Letting

$$\epsilon = ik\delta \quad (\text{D.5})$$

Equation D.4 can be cast as a quadratic equation of the form

$$\lambda^2 + \lambda(b_1 + \epsilon b_2 + O(\epsilon^2)) + (\epsilon c_1 + \epsilon^2 c_2 + O(\epsilon^3)) = 0 \quad (\text{D.6})$$

whose roots are given by

$$\lambda_1 = -b_1 + O(\epsilon) \quad (\text{D.7})$$

$$\lambda_2 = -\epsilon \frac{c_1}{b_1} + \epsilon^2 \left[ -\frac{c_2}{b_1} + \frac{c_1 b_2}{b_1^2} - \frac{c_1^2}{b_1^3} \right] + O(\epsilon^3) \quad (\text{D.8})$$

with

$$b_1 = \alpha_1 + \beta_1 + \alpha_2 + \beta_2 \quad (\text{D.9})$$

$$b_2 = \alpha_3 + \alpha_4 - \beta_3 - \beta_4$$

$$\begin{aligned} c_1 &= \alpha_1\alpha_2 - \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 - \beta_3) \\ &\quad + (\beta_1 + \alpha_2)(\alpha_4 - \beta_4) \end{aligned}$$

$$c_2 = (\alpha_3 - \beta_3)(\alpha_4 - \beta_4) - \frac{\alpha_1\alpha_2 + \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 + \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 + \beta_4)}{2}$$

Thus two eigenvalues of matrix **A** are given by

$$\lambda_1 = -(\alpha_1 + \beta_1 + \alpha_2 + \beta_2) + O(k\delta) \quad (\text{D.10})$$

$$\lambda_2 = -ik\delta \left[ \frac{\alpha_1\alpha_2 - \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 - \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 - \beta_4)}{\alpha_1 + \beta_1 + \alpha_2 + \beta_2} \right] - (k\delta)^2 \left[ \frac{\alpha_1\alpha_2 + \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 + \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 + \beta_4) - 2(\alpha_3 - \beta_3)(\alpha_4 - \beta_4)}{2(\alpha_1 + \beta_1 + \alpha_2 + \beta_2)} - (\alpha_1\alpha_2 - \beta_1\beta_2 - (\alpha_1 + \beta_2)(\alpha_4 - \beta_4) - (\beta_1 + \alpha_2)(\alpha_3 - \beta_3)) \right] \times \frac{(\alpha_1\alpha_2 - \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 - \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 - \beta_4))}{(\alpha_1 + \beta_1 + \alpha_2 + \beta_2)^3} + O((k\delta)^3) \quad (\text{D.11})$$

The effective spatial diffusion,  $D$ , is computed from the decay rate of the long wave modes of  $\rho = \rho_1 + \rho_2$ . The eigenvector corresponding to the eigenvalue  $\lambda_1$  is

$$\begin{bmatrix} c_1 \\ c_2 \end{bmatrix} = \begin{bmatrix} 1 \\ -1 \end{bmatrix} + O(k\delta) \quad (\text{D.12})$$

The decay of this eigenmode governs the convergence to equilibrium between states  $P_n$  and  $P'_n$ . In the transition diagram of Fig. 3, the decay of the eigenmode (D.12) represents the relaxation to equilibrium in the reaction (vertical) direction, whereas the dispersion in the horizontal direction is the effective spatial “diffusion,” characterized by  $D$ .

Because the eigenmode (Eq. D.12) satisfies  $\rho = \rho_1 + \rho_2 \approx 0$ , its decay rate does not affect the magnitude of long wave modes of  $\rho = \rho_1 + \rho_2$ . Thus the eigenvalue  $\lambda_1$  is not related to the spatial effective diffusion  $D$ . The eigenvalue  $\lambda_2$  governs the evolution of long wave modes of  $\rho = \rho_1 + \rho_2$ . When the eigenvector corresponding to  $\lambda_2$  is taken as the initial value for  $[c_1, c_2]$  in Eq. D.1, the solution of Eq. 4 is given by

$$\begin{bmatrix} \rho_1(n, t) \\ \rho_2(n, t) \end{bmatrix} = \exp\left( ik\delta \left( n - \frac{-\Im(\lambda_2)}{ik\delta} t \right) - k^2 \left( \frac{-\Re(\lambda_2)}{k^2} t \right) \right) \cdot \begin{bmatrix} c_1(0) \\ c_2(0) \end{bmatrix} \quad (\text{D.13})$$

where  $\Im(\lambda_2)$  and  $\Re(\lambda_2)$  are the imaginary and real parts of the eigenvalue  $\lambda_2$ .

The imaginary part of eigenvalue  $\lambda_2$  is responsible for the propagation (average velocity) of long wave modes, and the real part of  $\lambda_2$  is responsible for the decaying (dissipation) of long wave modes.

From Eqs. D.11 and D.13, we recover the average velocity  $v$  (in bp/s) given by Eq. 5:

$$v = \frac{\alpha_1\alpha_2 - \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 - \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 - \beta_4)}{\alpha_1 + \beta_1 + \alpha_2 + \beta_2} \quad (\text{D.14})$$

as well as the effective diffusion constant  $D$  (in  $\text{nm}^2/\text{s}$ ), given by Eq. 6:

$$D = \delta^2 \left[ \frac{\alpha_1\alpha_2 + \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 + \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 + \beta_4) - 2(\alpha_3 - \beta_3)(\alpha_4 - \beta_4)}{2(\alpha_1 + \beta_1 + \alpha_2 + \beta_2)} \right] \quad (\text{D.15})$$

$$\begin{aligned} & - (\alpha_1\alpha_2 - \beta_1\beta_2 - (\alpha_1 + \beta_2)(\alpha_4 - \beta_4) \\ & - (\beta_1 + \alpha_2)(\alpha_3 - \beta_3)) \\ & \times \frac{(\alpha_1\alpha_2 - \beta_1\beta_2 + (\alpha_1 + \beta_2)(\alpha_3 - \beta_3) + (\beta_1 + \alpha_2)(\alpha_4 - \beta_4))}{(\alpha_1 + \beta_1 + \alpha_2 + \beta_2)^3} \end{aligned}$$

*The randomness parameter and the effective diffusion constant*

We have defined the effective diffusion constant,  $D$ , for long times as

$$D = \frac{\text{var}(x(t))}{2t} \quad (\text{D.16})$$

Thus  $D$  accounts for both variance due to Brownian diffusion in space and the randomness along the reaction coordinate. For the case of negligible kinetic dispersion,  $\text{var}(x(t)) = (2t \times \text{Brownian diffusion constant})$ . So the effective diffusion can be regarded as a generalization that accounts for both spatial and kinetic dispersion. This effective diffusion constant is related to the “randomness parameter” defined by Schnitzer and Block in their analysis of the statistics of kinesin progression (Schnitzer and Block, 1995). Their randomness parameter,  $r$ , is defined as

$$r = \lim_{t \rightarrow \infty} \frac{\text{var}(x(t))}{\delta \langle x(t) \rangle} \quad (\text{D.17})$$

So the randomness parameter  $r$  and the effective diffusion constant  $D$  are related by

$$r = \frac{2D}{\delta v} \quad \text{or} \quad D = \frac{\delta v}{2} r \quad (\text{D.18})$$

Both the randomness parameter,  $r$ , and the effective diffusion,  $D$ , are characterized by the variance of the position  $x(t)$ , and contain the same information. Any statements in terms of  $r$  can be translated to ones in terms of  $D$  and vice versa. However, it is more appropriate to use the randomness parameter  $r$  in some situations, whereas the effective diffusion constant  $D$  is more appropriate in others. For example, suppose that the motion of a particle is described by steps of size  $\delta$ , and each step involves a sequence of reaction processes. Suppose that the reaction processes can be reversible but the spatial steps are not (i.e., the motor only goes forward). In this case,  $1/r$  gives a lower bound on the number of reaction processes per step (i.e.,  $1/r < \text{number of reaction processes per step}$ ) (Schnitzer and Block, 1995). Equivalently,  $\delta v/2D$  gives the same information, but the randomness parameter  $r$  is more appropriate for this problem. Note that in this problem the thermal fluctuations do not affect the particle position,  $x(t)$ . If  $x(t)$  represents the position of an object that is connected by a spring to the particle—e.g., as in the kinesin model, where  $x(t)$  is the position of the bead connected by a spring to the kinesin head (Schnitzer and Block, 1995)—then the contribution of thermal fluctuations to the variance of  $x(t)$  is a constant, independent of time. Thus the variance of  $x(t)$  is essentially caused by the randomness of the reaction processes. So it is not surprising that this problem is better treated by using the randomness parameter,  $r$ .

Now consider the case in which both the reaction processes and the spatial steps are reversible. Alternatively, let every reaction process be reversible; in particular, the reaction process that leads from the end of the

current step to the beginning of the next step. In this case, the system could have positive, zero, or negative average velocity. Thus when the average velocity is very small,  $r$  is very large. It is even possible for  $r$  to have negative value. In this case, the number of reaction processes per step is not given by  $1/r$ . It is clear that when the backward spatial or reaction steps are allowed, neither  $r$  nor  $D$  gives information about the number of reaction processes per step.

Finally, consider the problem where the particle is driven by a constant force,  $f$ , within a periodic potential,  $\psi(x)$ , and at the same time is subject to Brownian diffusion. The Fokker-Planck equation for this situation is

$$\frac{\partial \rho}{\partial t} = D_B \frac{\partial^2 \rho}{\partial x^2} + \frac{\partial}{\partial x} [(\psi'(x) - f)\rho] \quad (\text{D.19})$$

The periodic potential restricts the motion of the particle and thus tends to reduce the variance of the particle position  $x(t)$ . In this case, the effective diffusion constant,  $D$ , contains the effect of the potential, and more importantly, gives us a simplified model equation for the problem:

$$\frac{\partial \rho}{\partial t} = -v \frac{\partial \rho}{\partial x} + D \frac{\partial^2 \rho}{\partial x^2} \quad (\text{D.20})$$

where  $v$  is the average velocity. The simplified model equation (Eq. D.20) is valid for describing the long-time behavior of the particle governed by the more complicated equation (Eq. D.19). Thus the effective diffusion constant  $D$  is more appropriate in this situation.

To sum up, the randomness parameter,  $r$ , and the effective diffusion constant,  $D$ , contain the same information about the model, but each is more appropriate in different situations. Neither one is more general than the other.

## E. Computation of rates from experiments

In this section, we calculate the transition rates ( $\alpha_i, \beta_i$ ),  $i = 1, 2, 3, 4$ , using the constraints obtained from the principle of detailed balance and from experimental measurements.

### Dependence of the transition rates on the load force and concentrations

The Markov chain shown in Fig. 3 is a projection of a more complete enumeration of the kinetic states shown in Fig. 10. Here the horizontal axis,  $n$ , is the length of the RNA transcript; transition rates that have a component in the horizontal direction (polymerization direction) are functions of the load force. The vertical axis,  $k$ , is the number of NTPs hydrolyzed from the initial time. The third axis represents the binary state of RNAP with either pyrophosphate bound (primed states), or no pyrophosphate bound (unprimed states). Some transition rates are also functions of concentrations of NTP,  $\text{PP}_i$ , or nucleotide monophosphate (NMP). As we will see, experimental measurements generally lead to constraints on numbers instead of on functions. Therefore, to determine the transition rates we will have to specify the form of their dependence on the load force and concentrations.

The polymerization by hydrolysis of NTP requires a gap of size  $\delta$  between the RNA tip and the barrier on RNAP. It also requires that an NTP be available from solution to bind to the substrate site on RNAP. Thus  $\alpha_1$  must be proportional to the concentration of NTP in solution and to the probability of having a gap of size  $\delta$ :

$$\alpha_1 = \alpha_{10}^0 \cdot [\text{NTP}] \cdot e^{-\omega} \quad (\text{E.1})$$

where  $\alpha_{10}^0$  is the association constant of NTP with RNAP, and  $\omega = f\delta/k_B T$ .

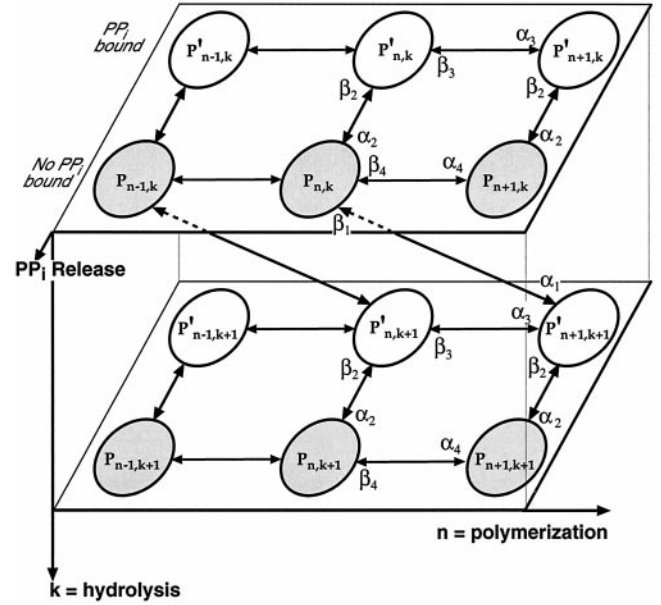


FIGURE 10 A more complete diagram of the chemical kinetic states where states with different free energies are shown as different states. For example, state  $P_{n,k+1}$  differs from state  $P_{n,k}$  in that one more NTP has been hydrolyzed by RNAP. Fig. 3 is a projection of this diagram. The horizontal axis,  $n$ , is the length of the RNA transcript; the vertical axis,  $k$ , is the number of NTPs hydrolyzed from the initial time; the third axis represents the binary state of RNAP with either pyrophosphate bound (primed states) or no pyrophosphate bound (unprimed states).

The rate of  $\text{PP}_i$  binding to RNAP (the reverse of  $\text{PP}_i$  release) is proportional to the concentration of  $\text{PP}_i$ :

$$\beta_2 = \beta_2^0 \cdot [\text{PP}_i] \quad (\text{E.2})$$

where  $\beta_2^0$  is the association constant of  $\text{PP}_i$  with RNAP.

It is thermodynamically possible for polymerization to take place directly from NTP that has hydrolyzed in solution ( $\alpha_3 \geq 0$ ). Although this rate is probably very small, it is necessary to retain it in the Markov chain in Fig. 10 because the reverse rate,  $\beta_3$ , is finite, and we need to ensure that the model obeys detailed balance (Hill, 1977, 1989). We show below that for direct NMP polymerization, we can neglect the force dependence, because polymerization by NMP is limited by the binding of NMP to RNAP. So the rates  $\alpha_3$  and  $\alpha_4$  are proportional to the concentration of NMP:

$$\alpha_3 = \alpha_3^0 \cdot [\text{NMP}] \quad (\text{E.3})$$

$$\alpha_4 = \alpha_4^0 \cdot [\text{NMP}] \quad (\text{E.4})$$

It follows from the principle of detailed balance that the depolymerization rates  $\beta_3$  and  $\beta_4$  have the form

$$\beta_3 = \beta_{30} \cdot e^{\omega} \quad (\text{E.5})$$

$$\beta_4 = \beta_{40} \cdot e^{\omega} \quad (\text{E.6})$$

Once the dependence on the load force and concentrations are specified, the eight transition rates ( $\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3, \alpha_4, \beta_4$ ) are totally determined by the eight numbers

$$\alpha_{10}^0, \beta_1, \alpha_2, \beta_2^0, \alpha_3^0, \beta_{30}, \alpha_4^0, \beta_{40} \quad (\text{E.7})$$

### Stall force

Now we express the stall force,  $f_s$ , in terms of these eight numbers. Setting the velocity to zero in Eq. 5, we have

$$\alpha_{10}^0 \alpha_2 [\text{NTP}] e^{-\omega} - \beta_1 \beta_2^0 [\text{PP}_i] + (\alpha_{10}^0 [\text{NTP}] e^{-\omega} + \beta_2^0 [\text{PP}_i]) \cdot (\alpha_3^0 [\text{NMP}] - \beta_{30} e^{\omega}) + (\alpha_2 + \beta_1) \cdot (\alpha_4^0 [\text{NMP}] - \beta_{40} e^{\omega}) = 0 \quad (\text{E.8})$$

Solving Eq. E.8 in terms of  $\exp(\omega)$ , we obtain the stall force  $f_s$ :

$$f_s = \frac{k_B T}{\delta} \ln(Q) \quad (\text{E.9})$$

where

$$Q = \frac{-q_1 + \sqrt{q_1^2 + 4q_2}}{2} \quad (\text{E.10})$$

$$q_1 = \frac{\alpha_{10}^0 \beta_{30} [\text{NTP}] + \beta_2^0 [\text{PP}_i] (\beta_1 - \alpha_3^0 [\text{NMP}]) - \alpha_4^0 [\text{NMP}] (\alpha_2 + \beta_1)}{\beta_{40} (\alpha_2 + \beta_1) + \beta_{30} \beta_2^0 [\text{PP}_i]} \quad (\text{E.11})$$

$$q_2 = \frac{\alpha_{10}^0 [\text{NTP}] (\alpha_2 + \alpha_3^0 [\text{NMP}])}{\beta_{40} (\alpha_2 + \beta_1) + \beta_{30} \beta_2^0 [\text{PP}_i]}$$

### Energy differences along the reaction paths

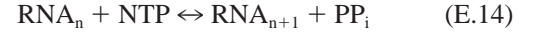
In Fig. 10, the path  $P_{n,k} \rightarrow P'_{n,k} \rightarrow P'_{n+1,k} \rightarrow P_{n+1,k} \rightarrow P_{n,k}$  forms a loop, so the free energy change along this path is zero. From the principle of detailed balance, we obtain that (Hill, 1977, 1989)

$$\left(\frac{\beta_2}{\alpha_2}\right) \cdot \left(\frac{\alpha_3}{\beta_3}\right) \cdot \left(\frac{\alpha_2}{\beta_2}\right) \cdot \left(\frac{\beta_4}{\alpha_4}\right) = 1 \quad (\text{E.12})$$

Equation E.12 yields our first constraint on the eight numbers listed in Eq. E.7:

$$\frac{\alpha_3^0}{\beta_{30}} = \frac{\alpha_4^0}{\beta_{40}} \quad (\text{E.13})$$

The path  $P_{n,k} \rightarrow P'_{n+1,k+1} \rightarrow P_{n+1,k+1}$  represents a typical transcription cycle where polymerization follows hydrolysis of NTP. In this cycle, an NTP is hydrolyzed, the nucleotide residue is added to the 3'-hydroxyl end of the nascent RNA chain, and finally the pyrophosphate is released. Schematically, the overall reaction of this transcription process is represented by (Erie et al., 1992)



The energy difference between the state  $P_{n,k}$  and the state  $P_{n+1,k+1}$  is then given by

$$G(P_{n+1,k+1}) - G(P_{n,k}) = \Delta G_{\text{NTP}, n \rightarrow n+1}^0 + k_B T \cdot \ln\left(\frac{[\text{PP}_i]}{[\text{NTP}]}\right) + f\delta \quad (\text{E.15})$$

where  $\Delta G_{\text{NTP}, n \rightarrow n+1}^0$  is the chemical free energy change of the reaction (Eq. E.14) at standard conditions (1 M  $\text{PP}_i$  and NTP, and pH 7).

The path  $P_{n,k} \rightarrow P_{n+1,k}$  represents a rare transcription cycle where an NMP from solution is added directly to the 3'-hydroxyl end of the nascent RNA chain. Schematically, the chemical reaction of this transcription process is represented by



The energy difference between the state  $P_{n,k}$  and the state  $P_{n+1,k}$  is then given by

$$G(P_{n+1,k}) - G(P_{n,k}) = \Delta G_{\text{NMP}, n \rightarrow n+1}^0 + k_B T \cdot \ln\left(\frac{1}{[\text{NMP}]}\right) + f\delta \quad (\text{E.17})$$

where  $\Delta G_{\text{NMP}, n \rightarrow n+1}^0$  is the chemical free energy change of the reaction (Eq. E.16) at standard conditions.

### Using experimental results

Table 3 lists the experimental results we will use to compute the transition rates. These experimental results give us five more constraints on the eight numbers listed in Eq. E.7.

**TABLE 3** Experimental results on RNAP transcription

Experiments	Results	Reference
1. Formation of poly (rA) from rATP, catalyzed by <i>E. coli</i> RNA polymerase	$\Delta G_{\text{NTP}, n \rightarrow n+1}^0$ $\approx -3.1$ kcal/mol $\approx -5.0 k_B T$	Erie et al. (1992); Kato et al. (1967); Peller (1976)
2. Formation of helical structure between two adjacent adenine residues in a homopolymer	$\Delta G_{\text{NMP}, n \rightarrow n+1}^0$ $\approx -1.0$ kcal/mol $\approx -1.6 k_B T$	Brahms et al. (1966); Erie et al. (1992)
3. RNAP transcription against an applied force (by an optical trap)	At concentrations $[\text{NTP}] = 1$ mM and $[\text{PP}_i] = 1$ $\mu\text{M}$ , the velocity at zero load is $\sim 30$ bp/s.	M. D. Wang, personal communication
4. RNAP transcription against an applied force (by an optical trap)	At concentrations $[\text{NTP}] = 1$ mM and $[\text{PP}_i] = 1$ mM, the velocity at zero load is $\sim 15$ bp/s.	M. D. Wang, personal communication; Yin et al. (1995)
5. RNAP transcription against an applied force (by an optical trap)	At concentrations $[\text{NTP}] = 1$ mM and $[\text{PP}_i] = 1$ $\mu\text{M}$ , the stall force is $f_s = 25\text{--}30$ pN.	M. D. Wang, personal communication

1. The equation obtained from the energy difference between  $P_{n,k}$  and  $P_{n+1,k+1}$  is

$$\left(\frac{\alpha_1}{\beta_1}\right) \cdot \left(\frac{\alpha_2}{\beta_2}\right) = \exp\left\{\frac{G(P_{n,k}) - G(P_{n+1,k+1})}{k_B T}\right\} \quad (\text{E.18})$$

which, from Table 3, leads to

$$\frac{\alpha_{10}^0 \alpha_2}{\beta_1 \beta_2^0} = \exp(5) \quad (\text{E.19})$$

2. The equation obtained from the energy difference between  $P_{n,k}$  and  $P_{n+1,k}$  is

$$\frac{\alpha_4}{\beta_4} = \exp\left\{\frac{G(P_{n,k}) - G(P_{n+1,k})}{k_B T}\right\} \quad (\text{E.20})$$

which leads to

$$\frac{\alpha_{40}^0}{\beta_{40}} = \exp(1.6) \quad (\text{E.21})$$

3. At concentrations  $[\text{NTP}] = 1 \text{ mM}$  and  $[\text{PP}_i] = 1 \mu\text{M}$  (the concentration of NMP is negligible), the maximum (no load) velocity  $v_M = 30 \text{ bp/s}$ :

$$\frac{\alpha_{10}^0 \alpha_2 10^{-3} - \beta_1 \beta_2^0 10^{-6} - \beta_{30}}{(\alpha_{10}^0 10^{-3} + \beta_2^0 10^{-6}) - \beta_{40}(\alpha_2 + \beta_1)} = 30 \quad (\text{E.22})$$

4. At concentrations  $[\text{NTP}] = 1 \text{ mM}$  and  $[\text{PP}_i] = 1 \text{ mM}$  (the concentration of NMP is negligible), the maximum (no load) velocity  $v_M = 15 \text{ bp/s}$ :

$$\frac{\alpha_{10}^0 \alpha_2 10^{-3} - \beta_1 \beta_2^0 10^{-3} - \beta_{30}}{(\alpha_{10}^0 10^{-3} + \beta_2^0 10^{-3}) - \beta_{40}(\alpha_2 + \beta_1)} = 15 \quad (\text{E.23})$$

5. At concentrations  $[\text{NTP}] = 1 \text{ mM}$  and  $[\text{PP}_i] = 1 \mu\text{M}$  (the concentration of NMP is negligible), the stall force is 25–30 pN. Let us take  $f_s = 28 \text{ pN}$ :

$$\frac{k_B T}{\delta} \ln(Q) = 28 \quad (\text{E.24})$$

where  $Q$  is given by Eqs. E.10 and E.11.

We now have six equations for the eight numbers listed in Eq. E.7. To close the set of equations, we make the following two assumptions, the validity of which must be determined experimentally:

1. We assume that the association constant of NTP to RNAP is similar to other nucleotide binding affinities. For example, the association constant of ATP to the catalytic sites on F1F0-ATP synthase is  $\sim 10^6 \text{ M}^{-1} \text{ s}^{-1}$  (Weber and Senior, 1997; Penefsky, 1991). Therefore, we shall take

$$\alpha_{10}^0 = 10^6 \quad (\text{E.25})$$

2. It is reasonable to assume that pyrophosphate, when bound on RNAP, interferes with direct polymerization by NMP from solution; that is,  $\text{PP}_i$  bound to RNAP increases the energy barrier of polymerization by NMP. This implies that the polymerization rate  $\alpha_3$  is much smaller than  $\alpha_4$ . More specifically, we assume that  $\text{PP}_i$  bound to RNAP raises the energy barrier of polymerization by the amount  $\Delta E_B$ ; we shall take  $\Delta E_B = 5k_B T$ :

$$\alpha_3^0 = \alpha_4^0 \cdot e^{-5} \quad (\text{E.26})$$

Fig. 11 shows that the model is not sensitive to the choice of the energy barrier increment  $\Delta E_B$ .

The complete set of equations consists of Eqs. E.13, E.19, E.21, E.22, E.23, E.24, E.25, and E.26. From these eight equations, we solve for the eight numbers listed in Eq. E.7. The transition rates are shown in Table 1.

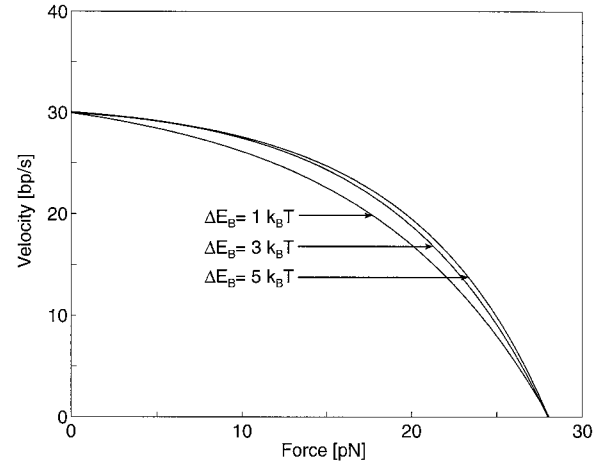


FIGURE 11 Load-velocity curves for different values of the energy barrier increment  $\Delta E_B$  at 1 mM NTP and 1  $\mu\text{M}$   $\text{PP}_i$ . The energy barrier increment  $\Delta E_B$  is caused by the presence of pyrophosphate after hydrolysis. This plot indicates that the model is not sensitive to the choice of  $\Delta E_B$ .

#### Justification for neglecting the force dependence in Eqs. E.3 and E.4

In Eqs. E.3 and E.4, we said that the rate of polymerization by NMP is proportional to the concentration of NMP and is independent of the load force. This appears to contradict Eq. E.1, where the rate of polymerization by hydrolysis of NTP is proportional to the concentration of NTP in solution and proportional to the probability of a gap of size  $\delta$  between the RNA tip and the barrier.

Here we show that this difference can be reconciled by the different time scales of the reaction processes. Suppose the polymerization process by NMP consists of two steps. First an NMP binds to a site that is very close to the RNA tip. Then it forms a bond with the RNA tip, elongating the RNA transcript by 1 bp. The polymerization process by NMP can be modeled by the Markov chain shown in Fig. 12 *a*. In this diagram,  $a_1$  is the rate of NMP binding to the site;  $b_1$  is the rate of NMP dissociating from the site;  $a_2$  is the rate of polymerization by an NMP on the site;  $b_2$  is the rate of depolymerization with the NMP going back to the site:

$$a_1 = a_1^0 \cdot [\text{NMP}] \quad (\text{E.27})$$

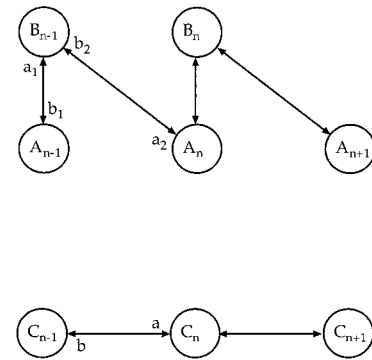


FIGURE 12 (a) Diagram of the Markov chain model of polymerization by NMP. First an NMP binds to a site that is very close to the RNA tip,  $A_{n-1} \rightarrow B_{n-1}$ . Then it forms a bond with the RNA tip, elongating the RNA transcript by 1 bp,  $B_{n-1} \rightarrow A_n$ . (b) Diagram of the simplified Markov chain model obtained from *a* by combining states  $B_{n-1}$  and  $A_n$  into a single state  $C_n$ . This simplification is justified if the transitions between states  $B_{n-1}$  and  $A_n$  are much faster than the transitions between states  $A_{n-1}$  and  $B_{n-1}$ .



$$a_2 = a_{20} \exp\left(\frac{-\eta f \delta}{k_B T}\right) \quad (\text{E.28})$$

$$b_2 = b_{20} \exp\left(\frac{(1-\eta)f\delta}{k_B T}\right) \quad (\text{E.29})$$

If the rate-limiting step in the polymerization process by NMP is the binding of NMP to the substrate site, then the transitions between states  $B_{n-1}$  and  $A_n$  are much faster than the transitions between states  $A_{n-1}$  and  $B_{n-1}$ . Hence states  $B_{n-1}$  and  $A_n$  can be combined into one state (call it  $C_n$ ). The Markov chain model in Fig. 12 *a* becomes the simplified one shown in Fig. 12 *b*.

The transition rates between  $C_{n-1}$  and  $C_n$  are given by

$$a = a_1 \cdot \frac{a_2}{a_2 + b_2}, \quad b = b_1 \cdot \frac{b_2}{a_2 + b_2} \quad (\text{E.30})$$

Now we assume that  $a_2 \gg b_2$ . That is, once the NMP binds to the site, the rate of polymerization by the NMP already on the site is large in comparison with the rate of depolymerization. Under this assumption, Eq. E.30 becomes

$$a = \text{const} \cdot [\text{NMP}], \quad b = \text{const} \cdot \exp\left(\frac{f\delta}{k_B T}\right) \quad (\text{E.31})$$

The authors thank Michelle Wang and Steve Block, who generously shared their unpublished data with us, and Peter von Hippel, who critically reviewed the manuscript and made many helpful suggestions.

H-YW was supported by a postdoctoral fellowship from the National Energy Research Scientific Computing Center, TE by postdoctoral support from Los Alamos National Laboratory, and GO by National Science Foundation grant DMS 9220719.

## REFERENCES

- Berg, H. C., and L. Turner. 1993. Torque generated by the flagellar motor of *Escherichia coli*. *Biophys. J.* 65:2201–2216.
- Brahms, J., A. M. Michelson, and K. E. van Holde. 1966. Adenylate oligomers in single- and double-strand conformation. *J. Mol. Biol.* 15:467–488.
- Chamberlin, M. J. 1994. New models for the mechanism of transcription elongation and its regulation. *Harvey Lect.* 88:1–21.
- Erie, D. A., T. D. Yager, and P. von Hippel. 1992. The single-nucleotide addition cycle in transcription: a biophysical and biochemical perspective. *Annu. Rev. Biophys. Biomol. Struct.* 21:379–415.
- Finer, J. T., R. M. Simmons, and J. A. Spudich. 1994. Single myosin molecule mechanics: piconewton forces and nanometre steps. *Nature.* 368:113–119.
- Hanggi, P., P. Talkner, and M. Borkovec. 1990. Reaction-rate theory: 50 years after Kramers. *Rev. Mod. Phys.* 62:254–341.
- Hill, T. L. 1977. Biochemical cycles and free energy transduction. *Trends Biochem. Sci.* 2:204–207.
- Hill, T. L. 1989. Free Energy Transduction and Biochemical Cycle Kinetics. Springer Verlag, New York.
- Hunt, A. J., F. Gittes, and J. Howard. 1994. The force exerted by a single kinesin molecule against a viscous load. *Biophys. J.* 67:766–781.
- Kato, K., J. M. Goncalves, G. E. Houts, and F. J. Bollum. 1967. Deoxynucleotide-polymerizing enzymes of calf thymus gland. II. Properties of the terminal deoxynucleotidyltransferase. *J. Biol. Chem.* 242:2780–2789.
- Krummel, B., and M. J. Chamberlin. 1992. Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase. Deoxyribonuclease I footprinting of defined complexes. *J. Mol. Biol.* 225:239–250.
- Landick, R. 1997. RNA polymerase slides home: pause and termination site recognition. *Cell.* 88:741–744.
- Mogilner, A., and G. Oster. 1996. Cell motility driven by actin polymerization. *Biophys. J.* 71:3030–3045.
- Molloy, J. E., J. E. Burns, J. Kendrick-Jones, R. T. Tregear, and D. C. White. 1995. Movement and force produced by a single myosin head. *Nature.* 378:209–212.
- Nudler, E., A. Mustaeiev, E. Lukhtanov, and A. Goldfarb. 1997. The RNA-DNA hybrid maintains the register of transcription by preventing backtracking of RNA polymerase. *Cell.* 89:33–41.
- Peller, L. 1976. On the free-energy changes in the synthesis and degradation of nucleic acids. *Biochemistry.* 15:141–146.
- Penefsky, H. S. 1991. Structure and mechanism of  $F_0F_1$ -type ATP synthases and ATPases. In *Advances In Enzymology*. A. Meister, editor. John Wiley and Sons, New York. 173–214.
- Peskin, C. S., G. M. Odell, and G. Oster. 1993. Cellular motions and thermal fluctuations: the Brownian ratchet. *Biophys. J.* 65:316–324.
- Peskin, C. S., and G. Oster. 1995. Coordinated hydrolysis explains the mechanical behavior of kinesin. *Biophys. J.* 68:202s–210s.
- Polyakov, A., E. Severinova, and S. A. Darst. 1995. Three-dimensional structure of *E. coli* core RNA polymerase: promoter binding and elongation conformations of the enzyme. *Cell.* 83:365–373.
- Risken, H. 1989. The Fokker-Planck Equation. Springer Verlag, New York.
- Schnitzer, M. J., and S. M. Block. 1995. Statistical kinetics of processive enzymes. *Cold Spring Harbor Symp. Quant. Biol.* 60:793–802.
- Schnitzer, M. J., and S. M. Block. 1997. Kinesin hydrolyses one ATP per 8-nm step. *Nature.* 388:386–390.
- Svoboda, K., and S. M. Block. 1994. Force and velocity measured for single kinesin molecules. *Cell.* 77:773–784.
- Svoboda, K., P. P. Mitra, and S. M. Block. 1994. Fluctuation analysis of motor protein movement and single enzyme kinetics. *Proc. Natl. Acad. Sci. USA.* 91:11782–11786.
- Weber, J., and A. E. Senior. 1997. Catalytic mechanism of  $F_1$ -ATPase. *Biochim. Biophys. Acta.* 1319:19–58.
- Yager, T. D., and P. von Hippel. 1987. Transcript elongation and termination in *Escherichia coli*. In *Escherichia coli and Salmonella typhimurium*. F. Neidhardt, editor. American Society for Microbiology, Washington, DC. 1241–1275.
- Yin, H., M. D. Wang, K. Svoboda, R. Landick, S. M. Block, and J. Gelles. 1995. Transcription against an applied force. *Science.* 270:1653–1657.