Motion of RNA Polymerase along DNA: A Stochastic Model

Frank Jülicher* and Robijn Bruinsma#
*Institut Curie, Physicochimie Curie, Section de Recherche, 75231 Paris Cedex 05, France, and #Department of Physics and Astronomy, University of California, Los Angeles, California 90095 USA

ABSTRACT RNA polymerase is a key transcription enzyme that moves along a DNA double helix to polymerize an RNA transcript. Recent progress in micromechanical experiments permits quantitative studies of forces and motion generated by the enzyme. We present in this paper a chemical kinetics description of RNA polymerase motion. The model is based on a classical chemical kinetics description of polymerization reactions driven by a free energy gain that depends on forces applied externally at the catalytic site. The RNA polymerase controlled activation barrier of the reaction is assumed to be strongly dependent on inhibitory internal strains of the RNA polymerase molecule. The sequence sensitivity of RNA polymerase is described by a linear coupling between the height of the activation barrier and the local DNA sequence. Our model can simulate optical trap experiments and allows us to study the dynamics of chemically halted complexes that are important for footprinting studies. We find that the effective stall force is a sequence-dependent, statistical quantity, whose distribution depends on the observation time. The results are consistent with the experimental observations to date.

INTRODUCTION

RNA polymerase (RNAP) plays a key role in the transcription of the genetic information encoded in DNA by controlling the synthesis of RNA chains. During the elongation of an RNA chain, after the binding of the polymerase to a promoter sequence on the DNA, the polymerase slides along the DNA while maintaining a high level of stability against dissociation. During sliding, the polymerase performs work to overcome various sources of energy dissipation such as viscous drag due to the relative motion of polymerase and DNA (sliding and rotation), and work required to open up the transcription bubble of the double-stranded DNA. The free energy required for this work is provided by the polymerization reaction of the nascent RNA chain when a ribonucleoside triphosphate (NTP) is added to the RNA chain under release of phosphate (PP_i) (Erie et al., 1992; Lewis, 1994). This polymerization occurs at a catalytic site *C* of the enzyme, which furthermore has binding sites for DNA and the nascent RNA strand (see Fig. 1). Motion of RNAP is influenced by signals in the DNA sequence that is transcribed to RNA. Several types of such signals exist where motion slows down (pause) or stops reversibly for an extended time (arrest), or where the enzyme detaches from the DNA (termination) (Landick, 1997).

Recently, the force generated by polymerase during elongation has been measured by the so-called optical trap method (Yin et al., 1995). Polymerase could be reversibly stalled if an external force on the order of 14 pN were applied. The observed stalling forces showed a broad statistical distribution. Typical stalling forces were relatively independent of the PP_i concentration, whereas, on the other hand, the maximum transcription velocity did decrease significantly with the concentration of PP_i.

These studies demonstrate explicitly that RNAP falls in a larger class of enzymes that can generate forces and motion along one-dimensional structures. The observation in recent years of micromechanical properties of more conventional motor proteins such as kinesins, which move along microtubules, or myosins, which move along actin filaments, has stimulated the development of stochastic models to describe and understand their behavior (Leibler and Huse, 1993; Prost et al., 1994; Peskin et al., 1994; Peskin and Oster, 1995; Derényi and Vicsek, 1996; Duke and Leibler, 1996; Jülicher et al., 1997). A comparison of the general properties of RNAP with those of motor proteins such as kinesins reveals fundamental differences that reflect the very different purposes of the two types of enzymes: whereas motor proteins are used for fast active transport in cells, RNAP has to produce an RNA strand that is an exact copy of the DNA template. RNAP has very high processivity, and it is able to recognize specific DNA sequences to reliably terminate transcription at well-defined positions along the DNA. These differences suggest that a model of RNAP motion could require a qualitatively different approach.

First, from a general point of view, there is a difference in symmetry. Kinesins or myosins move along cytosceletal filaments that have a polarity that determines the direction of motion, whereas RNAP slides along a DNA double helix, which on average is a nonpolar structure that in principle allows for motion in two opposite directions. The actual direction of motion of the RNAP is determined by the way RNAP binds to DNA at the promoter site. A second difference concerns the size of individual stepping events. For kinesins, a characteristic stepping distance of 8 nm has been observed that corresponds to the periodicity of tubulin monomers along a microtubule (Svoboda et al., 1994;

Received for publication 29 May 1997 and in final form 24 November 1997.

Address reprint requests to Dr. Frank Jülicher, Physicochemie, Institut Curie, 11 rue Pierre et Marie Curie, 75231 Paris Cedex 5, France. Tel.: 33-1-42-34-64-76; Fax: 33-1-40-51-06-36; E-mail: julicher@curie.fr.

0006-3495/98/03/1169/17 $2.00

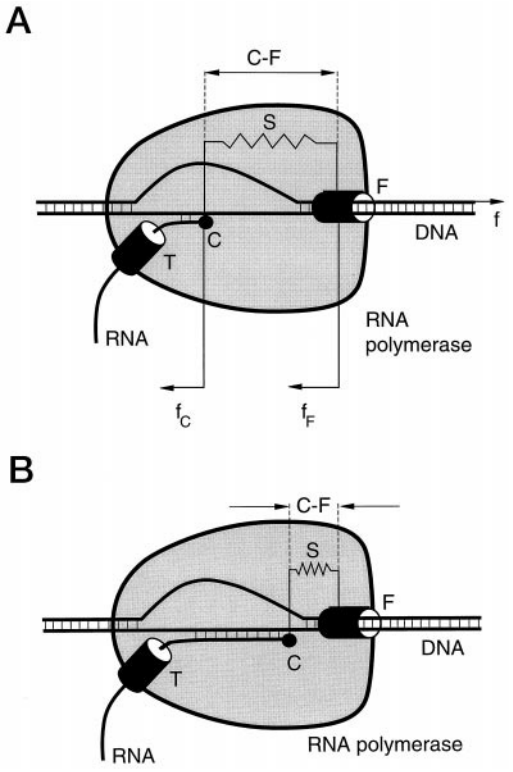

FIGURE 1 Schematic representation of RNAP moving along DNA. Within the enzyme, DNA is opened in a transcription bubble. At the catalytic site *C*, ribonucleotides are polymerized with a base sequence complementary to the template strand of DNA. The RNAP is bound to the DNA at a DNA binding domain in the front of the molecule (*F*), and the nascent RNA is threaded out to the exterior via a strong binding site *T*. Our model assumes an internal flexibility of the RNAP, represented by an elastic element *S*, and describes the motion of two degrees of freedom corresponding to the position of the catalytic site *C* and the distance $F - C$ between the front and the catalytic site. We distinguish the externally applied forces f_C and f_F acting on the catalytic site and the front, respectively. In an experimental situation, these forces are balanced by a force exerted on the DNA, $f = f_C + f_F$. (*A*) Relaxed state; (*B*) stressed state.

Schnitzer and Block, 1997; Hua et al., 1997). The corresponding step size for RNAP is the distance between two base pairs in a DNA double helix of $a = 0.34$ nm. This difference in step size leads to a difference in velocities and forces: RNAP moves more slowly by at least a factor of 10 than, for instance, kinesins. On the other hand, force levels are increased significantly: the observed forces generated by RNAP of at least 14 pN (Yin et al., 1995) can be compared to a maximum force of ~5 pN for kinesin (Meyhöfer and Howard, 1995). A second important difference concerns the way chemical energy is provided and consumed. RNAP motion is driven by the RNA polymerization reaction (Erie et al., 1992). The chemical energy is not provided by a certain concentration of available ATP, but rather by the "monomer"—a nucleoside triphosphate molecule—which is being added to the RNA chain. One polymerization step corresponds to one forward step. In the classical model of RNA polymerization, a stalled RNAP does not consume chemical energy and thus is in thermal equilibrium with the surrounding medium. For kinesin, on the other hand, the chemical reaction cycle is not directly coupled to motion: stalled kinesin can continue to consume ATP.

The fact that the step size of RNAP (0.34 nm) is small compared to the size of the molecule (~10–20 nm) suggests that RNAP does not move by a series of large steps (like kinesin). Instead, after a polymerization step, internal stresses within the enzyme produced by the polymerization must relax before the next polymerization step can take place. This stress relaxation leads to a sliding motion of RNAP over a distance on the order of 1 bp. These internal stresses are particularly important if we want to model the response to externally applied forces.

Conformational changes of enzymes involving internally "stressed" and "relaxed" states with different binding affinities for the substrate form a well-known general aspect of enzymatic action (Stryer, 1988). In the case of RNAP, there is direct evidence for structural flexibility and internal elasticity (Mustaev et al., 1993). For instance, the structural changes of the transcription complex during elongation can be monitored by the footprinting method, which measures the section of DNA protected by the RNAP against DNA-cutting enzymes (Metzger et al., 1989; Rice et al., 1991; Krummel and Chamberlin, 1992; Nudler et al., 1994; Wang et al., 1995; Chamberlin, 1995). The results of these studies indicate that RNAP indeed undergoes large, sequence-dependent changes during elongation. The interpretation of the footprinting studies is still under discussion, however (Landick, 1997). The so-called sliding clamp model that we will adopt in this paper assumes that the front-end DNA binding domain of RNAP (which we call *F* in the following) can act like a clamp that has to slide with respect to the DNA for motion to occur (Komissarova and Kashlev, 1997).

A difficulty of interpreting the results of footprint studies is that the precise microscopic origins of footprint size variations are not known. Significant variations in the footprint size of an elongating transcription complex have been observed that typically occur in the vicinity of pause, arrest, or termination sites and seem to be determined by the DNA sequence (Nudler et al., 1995; Chamberlin, 1995; Wang et al., 1995). These observations indicate the existence of structural changes of elongation complexes halted chemically at different positions along the DNA. An example is the possible detachment and sliding of the catalytic site with respect to the DNA and transcript (Reeder and Hawley, 1996; Komissarova and Kashlev, 1997). Elastic deformations of the enzyme could also lead to variations in footprint sizes (Chamberlin, 1995). Other phenomena, such as the formation of RNA secondary structures, could also influence footprint sizes.

It is the aim of this paper to provide a chemical kinetics description of the elongation of RNA chains under an applied force that takes into account the coupling between the force and the internal conformations of the enzyme. A chemical kinetics description of enzymatic polymerization has the advantage that it relies mostly on general principles

of nonequilibrium thermodynamics, albeit at the price of having to introduce rate constants that have to be determined experimentally. (Chemical kinetics descriptions of DNA polymerase are already used to analyze gel assays of T_4 polymerase (Creighton and Goodman, 1995).) We will consider two variants of our model with different force-transduction kinetics: 1) direct energy transduction and 2) energy transduction involving strain-dependent activation. The actual transduction mechanism is likely to be an intermediate between these cases. The model will be used to simulate numerically the sequence-dependent motion of RNAP along DNA and compare the results with optical trap and footprinting studies. A central question we aim to address concerns the "fidelity paradox." If we adopt the classical reaction scheme for RNA polymerization, then stalled RNAP complexes do not dissipate chemical energy unless there is "slippage," leading to transcription errors. On the other hand, the RNAP efficiency measured by Yip et al. indicates that stalled complexes in fact do dissipate a significant amount of chemical energy, yet the transcription fidelity of RNAP is very high under normal conditions.

In the next section we will discuss the classical chemical kinetics description of polymerization reactions (namely as a one-step Markov process) and apply it to elongation. This simplified model already gives insight into the different dependence of transcription velocity and stalling forces on the PP_i concentration. The sliding clamp picture of RNAP motion suggests that a one-step Markov model is not sufficient. A one-step model cannot account for the sequence dependence of elongation and for the observed broad distribution of stalling forces (Yin et al., 1995). In the third section we propose a generalized model that includes the internal strain as a second variable. The catalytic site proceeds according to a polymerization kinetics in single steps. This stepping leads to the buildup of internal strain, because the front F does not move immediately, because of "friction" between the clamp and the DNA. The kinetics of strain generation and relaxation require the introduction of an elastic energy of the polymerase molecule. The elastic modulus of the molecule as well as the strain relaxation rate are introduced as phenomenological parameters. The sequence sensitivity of motion in our model enters in a natural way through the activation barriers against strain relaxation. Our model in general allows for sequence dependence of both the RNA and DNA binding sites.

In the fourth section we examine the properties of our generalized model for the case without DNA sequence dependence by describing a mean-field theory together with Monte Carlo (MC) simulations. This analysis reveals that in addition to the thermodynamic stall force f_{stall}, a second critical value of the force can be defined. At this value (f_s) the velocity decreases exponentially and only slow "creep" motion remains. The force f_s is close to what in many practical conditions would be called "stall force." In such experiments, the thermodynamic stalling force is not observable because of a limited observation time. Our proposed resolution of the fidelity paradox relies on the identification of the experimentally measured stall force with f_s rather than with the thermodynamic stall force f_{stall}. In general, f_s can be small compared to f_{stall}.

In the fifth section we study the effect of DNA sequences on the properties of motion. Using Monte Carlo simulations of our model for motion along a given base sequence, we find that motion progresses with a mean velocity, but occasionally slows down or pauses if some rare sequence patterns are encountered, a behavior reminiscent of experimentally observed pausing or arrest events. We have monitored numerically the progression of RNAP along a DNA sequence under "optical trap" conditions with a harmonic restoring force to simulate an actual measurement of the stalling force. We obtained a distribution of stalling forces with a mean value below the thermodynamic stalling force, yet without permitting any slippage, indicating that the fidelity paradox has been removed. We then simulated RNAP progression under "footprinting" conditions by halting the RNAP at different base pairs. We did not find any large variations in the size of the RNAP during the elongation but after the RNAP had been brought to a halt, large, sequence-dependent size variations were observed after the relaxation that follows halting, consistent with the footprinting experiments. Our results appear to suggest that the observed large variation in footprint size should not be interpreted as evidence of large variations in the RNAP size during transcription.

POLYMERIZATION KINETICS

We will describe the polymerization of the nascent RNA by the integer variable n, which is the number of bases added to the RNA chain. This variable can also be interpreted as the position of the catalytic site along the DNA measured in bases from the initiation site of the transcription process. In the classical model of polymerization kinetics, it is assumed that the reaction rates k_+ and k_- of the polymerization step $n \rightleftharpoons n+1$ are independent of the polymerization index n. From the general principle of detailed balance, it then follows that

$$\frac{k^+}{k^-} = \exp(\Delta G/k_{\mathrm{B}}T), \tag{1}$$

with the free energy gain ΔG per polymerization step. For polymerization of an RNA chain by the addition of an NTP unit to the chain under release of PP_i, $\Delta G = \Delta\mu$ with

$$\Delta\mu = \Delta G_0 + k_{\mathrm{B}}T \ln\left(\frac{[\mathrm{NTP}]}{[\mathrm{PP_i}]}\right). \tag{2}$$

Here ΔG_0 is the standard free energy of the reaction. In general, ΔG_0 will be different for the four different monomers (A, U, G, and C) to be polymerized, and it can depend on the sequence of bases in the RNA chain that has already been generated. We will assume that this base dependence is not the dominant origin of sequence sensitivity and that $\Delta G_0 \simeq 3$ kcal/mol $\simeq 5k_{\mathrm{B}}T$ (Erie et al., 1992). Typical values

for $\Delta\mu$ are on the order of 5–12k_BT, depending on nucleotide concentrations.

If the polymerization is taking place under the presence of an external force f_C (the index C indicates that it is acting on the catalytic site) that opposes polymerization, we have

$$\Delta G = \Delta\mu - f_C a, \tag{3}$$

where a is the distance between bases along the DNA strand.

Only the ratio of forward and backward rates is fixed by the principle of detailed balance. To compute average velocities, we must specify the rate constants k^+ and k^- individually, which requires assumptions on the polymerization mechanism. In the classical model of polymerization of a chain (see, for instance, Oosawa and Asakura, 1975),

$$k^+ = k_0[\mathrm{NTP}]\exp(-\Delta U/k_BT), \tag{4}$$

where k_0 is a parameter that is independent of the concentration [NTP] and ΔU is the activation barrier for the reaction $n \rightarrow n + 1$. This activation barrier depends on the specific aspects of the chemical reaction. Note that ΔU can depend on the bias force f_C. Using Eq. 1, we obtain

$$\begin{aligned} k^- &= \tilde{k}_0[\mathrm{NTP}]\exp(-[\Delta\mu - f_C a]/k_BT) \\ &= \tilde{k}_0[\mathrm{PP_i}]\exp(-[\Delta G_0 - f_C a]/k_BT), \end{aligned} \tag{5}$$

where $\tilde{k}_0 = k_0 e^{-\Delta U/k_BT}$.

The quantities of interest that can be observed are the average velocity,

$$v = (k^+ - k^-)a, \tag{6}$$

and the "diffusion coefficient,"

$$D = (k^+ + k^-)/2a^2, \tag{7}$$

which characterizes the variance of motion (Svoboda et al., 1994; Schnitzer and Block, 1997). Using Eqs. 4 and 5, one finds that

$$v = a\tilde{k}_0([\mathrm{NTP}] - [\mathrm{PP_i}]\exp(-[\Delta G_0 - f_C a]/k_BT)), \tag{8}$$

and

$$D = \frac{\tilde{k}_0}{2a^2}([\mathrm{NTP}] + [\mathrm{PP_i}]\exp(-[\Delta G_0 - f_C a]/k_BT)). \tag{9}$$

If $\Delta G_0 - f_c a$ is positive and large compared to k_BT, the backward rate k^- is small. In this case the dimensionless parameter $2D/va$ is of order 1, which is typical for a biased random process obeying Poissonian statistics (Svoboda et al., 1994). If, however, $\Delta G_0 - f_c a$ is a large negative number, then $k^+ \ll k^-$ and backward motion occurs. From Eq. 8 it follows that the stalling force is given by

$$f_{\mathrm{stall}} = \frac{\Delta\mu}{a} = \frac{\Delta G_0}{a} + \frac{k_BT}{a}\ln\left(\frac{[\mathrm{NTP}]}{[\mathrm{PP_i}]}\right). \tag{10}$$

We will refer to f_{stall} as the "thermodynamic" stalling force. Note that the stalling force is only weakly dependent on the concentrations [NTP] and [PP_i], whereas the average velocity depends linearly on [NTP] and [PP_i] according to Eq. 8 as observed by Yin et al. (1995). Our argument shows that this observation is consistent with a chemical kinetics picture of polymerization. There is, however, a basic difficulty with Eq. 10. Using a typical value $\Delta\mu = 10k_BT$ and $a \simeq 3.4$ Å, one obtains $f_{\mathrm{stall}} \simeq 100$ pN, which is significantly larger than the experimentally obtained stall forces of ~14 pN (Yin et al., 1995).

In obtaining expression in Eq. 8 for the velocity, we assumed implicitly that the activation barrier did not depend on the applied force f_C. We will call this situation "direct energy transduction" and refer to it as model A. Physically, model A assumes that the activated intermediate state of the chemical reaction $n \rightarrow n + 1$ (i.e., the addition of one NTP) involves no internal strain or displacement of the polymerase that couples to the applied force. If the intermediate state does involve internal strain coupled to the applied force, energy transduction will be strain dependent. Assume, for example, the most extreme case, with $\Delta U = \Delta U_0 + f_C a$, which we call model B. Physically, model B describes an intermediate state for the $n \rightarrow n + 1$ reaction, which, as a result of a thermal fluctuation, already has moved forward by 1 bp before the next NTP is put in position. The polymerization step then "freezes" the shifted enzyme in its new position. A scenario of this type, which resembles a Brownian ratchet, was proposed by Yager and von Hippel (1987). In this case,

$$v = k_0[\mathrm{NTP}]e^{-\Delta U_0/k_BT}(e^{-f_C a/k_BT} - e^{-f_{\mathrm{stall}} a/k_BT}). \tag{11}$$

If $f_{\mathrm{stall}} \gg k_BT/a$, the observed velocity vanishes exponentially as a function of the external force f_C and is almost zero when f_C is large compared to k_BT/a. For $f_C \gg k_BT/a$, motion may not be observable within realistic observation times, which could lead to the observation of "apparent" stalling forces of $f_s \simeq k_BT/a \simeq 10$ pN. Note that this apparent stalling force f_s does not depend on nucleotide concentrations. This interpretation assumes that the backward reaction is negligibly small. The stall force predicted by model B appears to be in better agreement with the measured stall forces. However, there is a basic problem with model B as well. The transcription velocity (Eq. 11) of model B is not consistent with the experiments of Yin et al., because they detect a dependence of [PP_i] on the observed transcription velocity. We can conclude that neither model A nor model B is consistent with the experimental observations.

Efficiency and fidelity

A key quantity of interest for the following is the "efficiency," defined for $f \leq f_{\text{stall}}$ as

$$\eta \equiv f_C a/\Delta\mu = f_C/f_{\text{stall}}. \tag{12}$$

It is the ratio of mechanical work performed per unit time by the RNAP against the external force, divided by the consumption rate of chemical energy. If the chemical energy for the forward motion is obtained from a simple polymerization reaction, then the energy consumption rate is necessarily equal to the velocity of transcription times the free energy gain $\Delta\mu$ per step divided by the step size a. Each monomer that has been taken out of solution and added to the polymer chain contributes one step to the motion that consumes $\Delta\mu$ in free energy. Under stalling conditions, the work done $f_C a$ per step against the external force must equal $\Delta\mu/a$, and the energy consumption rate must vanish. In this simple scheme, a stalled complex is thus in thermodynamic equilibrium while $\eta = 1$ at stalling.

The validity of this scenario depends on the assumption that no other process consumes chemical energy, apart from the forward motion of the catalytic site. If a stalled RNAP would continue to copy repeatedly the same DNA section over and over again ("slippage") with the catalytic site moving forward and backward, then the stalled complex would continue to consume energy and the efficiency would be less than that under stalling conditions. This would imply, however, that the RNA generated is not a copy of the original DNA strand. Another scenario for obtaining efficiencies of less than 1 is that in which the standard reaction scheme for polymerization (Erie et al., 1992),

$$\text{RNA}_n + \text{NTP} \rightleftharpoons \text{RNA}_{n+1} + \text{PP}_i, \tag{13}$$

must be corrected. We will assume in this paper that slippage does not occur, and that the standard polymerization reaction scheme is correct. This means that the efficiency should be one under stalling conditions, in disagreement with the observations of Yin et al. (1995). This is the "fidelity paradox" mentioned in the Introduction. Note that model B avoids the difficulty, because the apparent stalling force $k_B T/a$ can be much less than the thermodynamic stalling force f_{stall}. We saw, however, that model B cannot explain the dependence of the transcription velocity on the PP_i concentration.

We will describe in the next section a generalized Markov model that can solve some of the discrepancies between the observed behavior of stall forces and the naive polymerization model described above. In this generalized model, the velocity again decreases exponentially before the thermodynamic stall force is even reached. However, the typical force f_s for which this happens depends on the model parameters and, in particular, on the PP_i concentration.

GENERALIZED MARKOV MODEL

In the previous section we reviewed the chemical kinetics single-site description of chain polymerization. Although this description is able to explain a number of observations on RNA elongation, by its very nature it cannot address the sequence dependence and does not take into account the internal flexibility of the enzyme. In this section we will describe the motion of RNAP in terms of two (rather than one) stochastic variables. The first is the integer index n, which, just as in the previous section, represents the position of the catalytic site C along the DNA. Alternatively, we can consider n as the length of the RNA chain produced so far. The second stochastic variable is the integer m, which measures the internal deformation of the polymerase during elongation. Each forward step of the catalytic site, with the position of the clamp F kept fixed, corresponds to an increase to $m + 1$ and leads to the buildup of internal strain. At the same time, the distance $C - F$, between F and the catalytic site C, is decreased by one step (see Fig. 1). Note that m assumes only positive values. Studies testing the structural deformability have shown that internal flexibility allows for deformations on the order of 8 bp (Mustaev et al., 1993); thus we assume m to vary somewhere between 0 and ~ 8.

To model an optical trap experiment where external forces are applied to the enzyme, we assume that the polymerase is fixed on a support and a force f is exerted on the DNA. This induces a counterforce $-f$ exerted in the opposite direction by the support. The applied force

$$f = f_C + f_F \tag{14}$$

is chosen to be positive if it opposes motion. It is divided into two contributions, f_C and f_F: the fraction f_C acts on the catalytic site, and the fraction f_F acts at the front (see Fig. 1). The exact way in which the total force is divided between f_C and f_F in principle depends on the details of the molecular shape, its elastic properties, and the details of the attachment to the substrate. We will therefore consider f_C and f_F in the following as two independent variables.

We will now construct a chemical kinetics description for the pair of variables (n, m) but restrict ourselves to the case of uniform DNA (sequence dependence will be discussed later). The kinetics of the polymerase motion will be assumed to consist of repeats of two types of elementary processes: step elongation and step relaxation. These two processes will be treated as chemical reactions. The first process is then

$$\text{(I)} \quad (n, m) \underset{-}{\overset{+}{\rightleftharpoons}} (n + 1, m + 1). \tag{15}$$

One NTP is added (+) or removed (−) from the RNA chain, the position of the catalytic site C changes by 1, and the position of the front F is fixed. The strain variable m increases (respectively decreases) by 1. The forward rate will be denoted by $k^+(m)$, the backward rate by $k^-(m + 1)$. The second process is

$$\text{(II)} \quad (n, m) \underset{+}{\overset{-}{\rightleftharpoons}} (n, m - 1). \tag{16}$$

The deformation m of the enzyme decreases ($-$) (respectively increases ($+$)) by 1, and the position of the catalytic site C is fixed. This step could correspond to the threading of one RNA monomer through the tight binding site, whereas the ($+$) step is rare and plays no role. The rate ($-$) for strain relaxation will be denoted by $r^-(m)$, and the rate for strain loading by $r^+(m-1)$.

The ratios $k^+(m)/k^-(m+1)$ and $r^+(m)/r^-(m+1)$ are determined by the condition of detailed balance:

$$\frac{k^+(m)}{k^-(m+1)} = \exp(\Delta G_{\mathrm{I}}(m)/k_{\mathrm{B}}T) \tag{17}$$

$$\frac{r^+(m)}{r^-(m+1)} = \exp(\Delta G_{\mathrm{II}}(m)/k_{\mathrm{B}}T), \tag{18}$$

where $\Delta G_{\mathrm{I}}(m)$ and $\Delta G_{\mathrm{II}}(m)$ denote the free energy difference between the two states involved in the "reactions" I and II. To evaluate ΔG_{I} and ΔG_{II}, we introduce $\mathscr{E}(m)$, the internal elastic free energy of the polymerase molecule. The function $\mathscr{E}(m)$ describes the deformation energy cost of reducing the C–F distance and/or collecting additional RNA monomers in the loose binding site.

The free energy difference $\Delta G_{\mathrm{I}}(m)$ for polymerization step I is

$$\Delta G_{\mathrm{I}}(m) = \Delta\mu + \mathscr{E}(m) - \mathscr{E}(m+1) - af_{\mathrm{C}}, \tag{19}$$

with $\Delta\mu$ given by Eq. 2. The free energy difference $\Delta G_{\mathrm{II}}(m)$ for strain relaxation step II is

$$\Delta G_{\mathrm{II}}(m) = \Delta E + \mathscr{E}(m) - \mathscr{E}(m+1) + af_{\mathrm{F}}. \tag{20}$$

Here ΔE is the free energy cost of opening up the double-stranded DNA into two single strands within the polymerase by 1 bp (it is thus on the order of a hydrogen bond energy). Note the different sign of the forces in Eqs. 19 and 20: the retarding force hinders polymerization but assists strain loading.

To completely define the rates, we must again specify the energy barriers $\Delta U_{\mathrm{I}}(m)$ and $\Delta U_{\mathrm{II}}(m)$ for the two processes. First we discuss process I. We will distinguish the two models for activation discussed in the previous section: model A, with an activation barrier that is unaffected by the applied force, and model B, with the barrier increasing linearly with the applied force. For model A:

$$k^+(m) = k_0^+[NTP] \tag{21a}$$

$$k^-(m+1) = k^+(m)\exp(-\Delta G_{\mathrm{I}}(m)/k_{\mathrm{B}}T). \tag{21b}$$

(model A; $\Delta G_{\mathrm{I}} > 0$)

In model A we must allow for depolymerization if the retarding force is large, so that $f_{\mathrm{C}} > \Delta\mu/a$. In this case it is possible that $\Delta G_{\mathrm{I}} < 0$. We will assume that the depolymerization rate $k^-(m)$ is now controlled by the reaction rate of the $\mathrm{PP_i}$ with the RNA chain (as in the previous section), with $k^+(m)$ now given by detailed balance:

$$k^-(m) = k_0^-[\mathrm{PP_i}] \tag{22a}$$

$$k^+(m) = k^-(m+1)\exp(\Delta G_{\mathrm{I}}(m)/k_{\mathrm{B}}T) \tag{22b}$$

(model A; $\Delta G_{\mathrm{I}} < 0$)

In model B, we assume a strain-dependent polymerization: the polymerase in its ground state with catalytic site C at position n is unreactive. Thermal fluctuations allow the site to move forward and backward. As soon as a forward fluctuation to the neighboring site $n+1$ occurs, a reactive state is attained and the polymerization takes place, and C rests at $n+1$. Because such a forward fluctuation occurs with a probability $\sim \exp([\mathscr{E}(m+1) - \mathscr{E}(m) + af_{\mathrm{C}}/k_{\mathrm{B}}T)$, this mechanism corresponds in our model to a strain-dependent activation barrier,

$$\Delta U_{\mathrm{I}}(m) = \mathscr{E}(m+1) - \mathscr{E}(m) + af_{\mathrm{C}}. \tag{23}$$

Reduced substrate affinity of strained enzymes is in fact held to be an important feature of allosteric proteins in general (Monod et al., 1965). Therefore we obtain

$$k^+(m) = k_0^+[\mathrm{NTP}]\exp(-\Delta U_{\mathrm{I}}(m)/k_{\mathrm{B}}T) \tag{24a}$$

$$k^-(m+1) = k^+(m)\exp(-\Delta G_{\mathrm{I}}(m)/k_{\mathrm{B}}T). \tag{24b}$$

(model B)

In model B we can continue to use Eq. 24, even for $\Delta G_{\mathrm{I}} < 0$, provided $-\Delta G_{\mathrm{I}}(m) < \Delta U_{\mathrm{I}}(m)$.

We now want to turn to process II: the sliding of the front F and/or the threading of RNAP (with rate r^-) through the tight binding site while the catalytic site stays fixed. The energy barrier for this process, ΔU_{II}, contains three parts:

$$\Delta U_{\mathrm{II}}(m) = \Lambda - \frac{1}{2}[\mathscr{E}(m+1) - \mathscr{E}(m)] + \frac{1}{2}af_{\mathrm{F}}. \tag{25}$$

The first term Λ, is a constant independent of m and the applied force. It represents the energy barrier to sliding the site F by one step, threading one RNA monomer through the tight binding site while assuming $f_{\mathrm{F}} = 0$, and neglecting the elastic force between C and F. The second term, half the energy gain of strain relaxation by one step, represents an estimate of the lowering of the activation barrier as a result of elastic strain. The last term represents the increase of the activation energy barrier due to the retarding force f_{F} exerted on the site F. It is possible to derive Eq. 25, assuming that the motion of site F along the DNA can be represented as a periodic potential with period a. The choice of a sawtooth potential leads to Eq. 25. Different periodic potentials give similar results, albeit with different numerical prefactors.

The reaction rates of step II consistent with detailed balance are

$$r^-(m+1) = r_0\exp(-\Delta U_{\mathrm{II}}(m)/k_{\mathrm{B}}T) \tag{26a}$$

$$r^+(m) = r^-(m+1)\exp(\Delta G_{\mathrm{II}}(m)/k_{\mathrm{B}}T), \tag{26b}$$

($\Delta U_{\mathrm{II}} > 0$)

provided $\Delta U_{\mathrm{II}} > 0$. For large m, the barrier for stress

relaxation decreases (see Eq. 25). If $\Delta U_{II}(m)$ turns negative, then stress relaxation is no longer activated: the F site can slide forward immediately. In this case,

$$r^-(m+1) = r_0 \tag{27a}$$

$$r^+(m) = r^-(m+1)\exp(\Delta G_{II}(m)/k_BT). \tag{27b}$$

$$(\Delta U_{II} < 0)$$

The main uncertainty in our model—apart from the value of Λ and the nature of the force-transduction—is the elastic free energy $\mathcal{E}(m)$ for $m = 0, 1, \ldots$. Although $\mathcal{E}(m)$ in principle could be measured, e.g., by the atomic force microscope, it is not a known function. We will assume that for $m = 0$ the polymerase is in a relaxed state, where the site F does not exert a force on the site C. For $m > 0$, we assume that $\mathcal{E}(m)$ increases monotonically. A simple m dependence for which these conditions hold is

$$\mathcal{E}(m) = \alpha m^2 \tag{28}$$

with an elastic constant α. In the following sections we will use Eq. 28. We have checked that our results are not sensitive to slight changes in the function $\mathcal{E}(m)$. If information on $\mathcal{E}(m)$ was available from experimental data, this could easily be incorporated into our model.

We now have fully specified our model for the motion of RNAP. In the next section we study this model, using a mean-field theory and Monte Carlo (MC) simulations.

MOTION ALONG HOMOGENEOUS DNA

During its progression along the DNA chain, the position of the catalytic site n undergoes statistical fluctuations around the average position $\langle n \rangle$, while the distance C–F varies around its mean value $l - \langle m \rangle$, where l denotes the distance between C and F in the relaxed state. In this section we are restricting ourselves to the case of homogeneous DNA for which the intrinsically random nature of chemical kinetics is the only source of statistical uncertainty. The aim of this section is to demonstrate that the objections raised against the simple polymerization model of the second section are indeed removed for the generalized Markov model presented in the previous section.

In principle, homogeneous segments of DNA (which are simply a repetition of a given base pair) can be used in experiments provided a promoter sequence is present. In general, however, DNA sequences are complex and influence the transcription process, as discussed in the Introduction. The structural heterogeneity of a typical DNA sequence introduces a second source of statistical fluctuations for both the position and for the C–F distance, which will be discussed in the next section. Here we present a "mean-field" theory for the averages $\langle n \rangle$ and $\langle m \rangle$, followed by MC simulations to examine the reliability of mean-field theory. Our mean-field theory gives physical insight into the parameter Λ introduced in Eq. 25, which will play a key role for the description of sequence dependence in the next section. The results of this section also help us to fix the values of the parameters α, k_0, and r_0.

Mean-field theory

For motion along homopolymer DNA, the system reaches a steady state with a well-defined distribution function $P(m)$ of the molecular deformation m. If this statistical distribution is narrowly peaked around some value, we can neglect m-fluctuations and use a mean-field theory where m is replaced by its average $\langle m \rangle$. To determine $\langle m \rangle$, we demand that for $\langle m \rangle$ to be time-independent, the forward rate $k^+(\langle m \rangle) + r^+(\langle m \rangle)$ for the increase in $\langle m \rangle$ must be equal to the reverse rate $k^-(\langle m \rangle + 1) + r^-(\langle m \rangle + 1)$ for the decrease in $\langle m \rangle + 1$. In other words,

$$k^+(\langle m \rangle) + r^+(\langle m \rangle) = k^-(\langle m \rangle + 1) + r^-(\langle m \rangle + 1). \tag{29}$$

This relation allows us to calculate $\langle m \rangle$ and the mean-field velocity $v = d\langle n \rangle/dt$ according to

$$\begin{aligned} v &= a(k^+(\langle m \rangle) - k^-(\langle m \rangle + 1)) \\ &= a(r^-(\langle m \rangle + 1) - r^+(\langle m \rangle)). \end{aligned} \tag{30}$$

Note that within mean-field theory we cannot calculate the variance of motion D in a meaningful way. This follows from the fact that D is determined by the fluctuations, most of which are neglected in mean-field approximation.

As described in detail in Appendix A, Eq. 29 leads to algebraic equations for $\langle m \rangle$ that are different for models A and B. A general result that holds for both models A and B is that thermodynamic stalling conditions are obtained if the total external force $f = f_C + f_F$ is equal to the stalling force,

$$f_{stall} = (\Delta\mu - \Delta E)/a. \tag{31}$$

In addition, the simple expression for the efficiency, $\eta = f/f_{stall}$ for $f < f_{stall}$, which we found for a simple polymerization model, holds for both models as well.

The mean-field predictions for the force-velocity curve of model A are as follows. The velocity v mainly depends on the total force, $f = f_C + f_F$. We can distinguish three qualitatively different situations corresponding to small, intermediate, and large external forces. For sufficiently weak forces $f \ll f_s$, where

$$f_s = (\Delta\mu - 2\Lambda)/a, \tag{32}$$

the velocity is given, to a good approximation, by $v = ar_0$. This is the RNAP velocity for $f = 0$, so the motion is not significantly affected by the retarding force in this regime.

As soon as $f \simeq f_s$, stress relaxation becomes activated and the velocity is reduced. The velocity decreases exponentially with the retarding force f,

$$v \simeq r_0 a \exp(a[f_s - f]/2k_BT) \tag{33}$$

in the intermediate range $f_s \ll f \ll f_{stall}$. The polymerase moves by a slow creep in this regime. For $f = f_{stall}$, motion

comes to a halt. Finally, in the regime of large external force with f exceeding f_{stall}, the direction of motion is reversed.

Note that the existence of the cross-over force f_s between the weak force regime with maximal velocity $v \simeq r_0 a$ and the intermediate force regime where the velocity decreases exponentially is qualitatively different from expressions (8) and (11), which we found for the simple polymerization model. From an experimental point of view, we may identify f_s as the applied force level for which motion slows down significantly and which for many practical cases is a good approximation for the observed stall force.

The efficiency for forces near this apparent stall force f_s is approximately on the order of $\eta = 1 - 2\Lambda/\Delta\mu$, which is less than 1. Provided we are permitted to make this identification, there is no fidelity paradox. Moreover, if we insert Eq. 2 for $\Delta\mu$ in Eq. 32, we find that this apparent stalling force is again only weakly dependent on the PP_i concentration, and it follows from Eq. 33 that the velocity v still decreases linearly with $[PP_i]$, as in Eq. 8. We thus can conclude that, at least within mean-field theory, the generalized Markov model has retained those features of model A discussed in the second Section that agree with the optical trap experiments, while it has resolved the fidelity problem.

If, for short observation periods, we do identify f_s as the experimentally measured stall force, then we can loosely interpret the quantity $2\Lambda/a$ in Eq. 32 as a microscopic analog of the static friction force. Note that for forces in the range $f_{stall} \gg f \gg f_s$, the transcription velocity is an exponentially sensitive function of the parameter Λ through the factor $\exp(af_s/2k_BT)$ (see Eq. 33). This sensitivity will play an important role for the sequence dependence discussed later. Using observed stalling forces of 14 pN to determine a value of Λ leads with $\Delta\mu \simeq 10k_BT$ to the estimate $2\Lambda/a \simeq 80$ pN and $\Lambda \simeq 4k_BT$. To see if this interpretation of the results is reasonable and if the mean-field theory applies, we will use numerical studies of our model.

Simulation results

The properties of our model can be studied by using the Monte Carlo simulation techniques described in Appendix B. This methods allows us to study the full behavior of our model, including the effects of fluctuations that have been neglected in mean-field theory. Moreover, it permits the study of sequence sensitivity as described in the next section. In our first example, we choose $\Delta\mu/k_BT = 10$, $\Lambda/k_BT = 6$, $\Delta E = 0$, $\alpha/k_BT = 0.4$, and $\bar{k}_0/r_0 = 2$. For these values, $f_s = 2k_BT/a$. Fig. 2 shows a simulation of model A: n increases roughly linearly with time t. The average compression in this case is $\langle m \rangle \simeq 12$, while $\delta m \simeq 2 \ll \langle m \rangle$, which indicates that a mean-field approximation is justified. For model B, we find with the same parameters a similar behavior. However, the velocity is decreased as a result of an extra activation barrier for polymerization. For model B, we observe a reduced average compression $\langle m \rangle \simeq 4$, because force transduction is less efficient. A mean-field theory in this case is less justified than for model A.

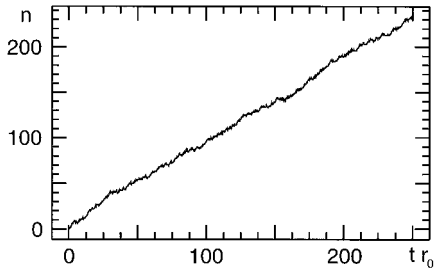

FIGURE 2 Polymerization index n versus time t for $\Delta\mu/k_BT = 10$ and $\Lambda/k_BT = 6$. This example uses the dynamics of model A.

Simulation results are displayed in Fig. 3 for both models A and B. We first discuss the simulations of model A. Fig. 3, *a* and *b*, shows velocities as a function of f_C for $f_F = 0$ and as a function of f_F for $f_C = 0$, respectively. The velocity saturates for negative forces (which assist motion) at a maximum value ar_0.

To compare our simulation results with mean-field theory, we first tested, for low force levels $f \ll f_{stall}$, whether the mean velocity v only depends on the total force $f = f_C + f_F$ as predicted by Eqs. 32 and 33. We indeed found that the dependency of v on f_C and on f_F was very similar (see Fig. 3, *a* and *b*). Moreover, as shown in Fig. 4, we found that for higher applied force levels, the average velocity decreased exponentially with f, as predicted by Eq. 33. On the other hand, for low force levels the velocity was relatively independent of f. As can be seen from Fig. 3, the cross-over between these two regimes occurs near the (negative!) apparent stall force f_s, in agreement with mean-field theory (we choose parameters such that f_s was negative, to enhance the contrast between f_s and f_{stall}). Looking at the behavior of $\langle m \rangle$ shown in Fig. 3, *c* and *d*, we note that $\langle m \rangle$ decreases linearly with increasing f_C, but that it is relatively independent of f_F. As shown in Fig. 3, *e* and *f*, δm is on the order of 1 over the whole range of f. This last result demonstrates that the basic validity condition of mean-field theory, namely that $\delta m/\langle m \rangle$ must be small compared to 1, is satisfied.

Our simulations of model B show a more pronounced exponential decay of the velocity, consistent with $v \approx \exp(-fa/k_BT)$, resulting from strain-dependent polymerization rates (see Fig. 3, *a* and *b*). Again, the levels of dependence of v on f_C and f_F are very similar, and the velocity saturates at the maximum value ar_0 for $f \ll 0$. Fig. 3, *c* and *d*, reveals that the average compression as a function of f_C is in general smaller for model B than for model A. In model B, the average compression now depends on the force f_F. For large f_F, this compression exceeds that measured for model A. The fluctuations δm are on the order of 1 in model B, independent of the external force (see Fig. 3, *e* and *f*).

Fig. 4 shows the influence of the energy barrier Λ for the case of model A. The average velocity v is shown as a function of the force f_C for different values of Λ. The exponential decay of velocity with force is well described by mean-field theory

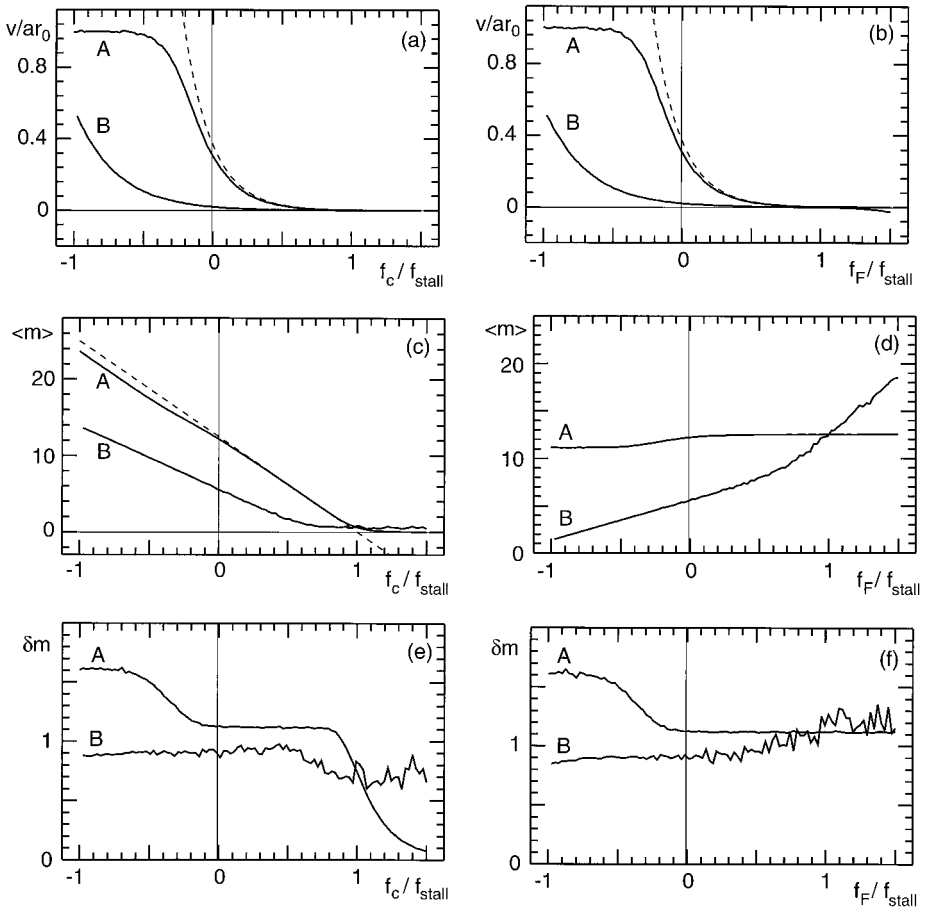

FIGURE 3 Simulation results for model A and model B for motion along homopolymer DNA. (*a*) Average velocity v as a function of the force f_C measured in units of the thermodynamic stalling force $f_{stall} = \Delta\mu/a$ acting at the catalytic site. Using $\Delta\mu = 10k_BT$ and $a = 3.4$ Å, one has, e.g., $f_{stall} \simeq 115$ pN. The broken line corresponds to Eq. 33. (*b*) Same diagram, but as a function of the force f_F acting at the front of the enzyme. (*c* and *d*) Average compression $\langle m \rangle$ as functions of f_C and f_F. The broken line in *c* corresponds to the linear approximation $(\Delta\mu - af_C)/2\alpha$ (see Eqs. A8 and A11). (*e* and *f*) Fluctuations of compression δm as functions of f_C and f_F.

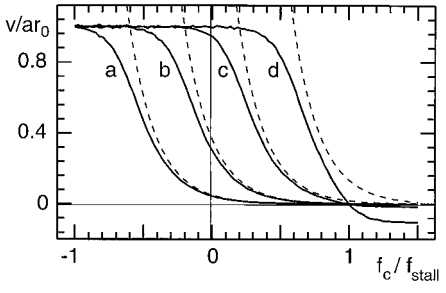

FIGURE 4 Average velocity v as a function of the force f_C for model A, measured in units of the thermodynamic stalling force f_{stall}, for $\Lambda/k_BT = 8$ (*a*), 6 (*b*), 4 (*c*), and 2 (*d*). The broken lines indicate the approximations as given by Eq. 33.

for $\Lambda/k_BT = 4$, 6, and 8 (Eq. 33). The figure demonstrates that for $\Lambda/k_BT < \Delta\mu/2$, where $f_s > 0$, the velocity saturates already for positive forces, whereas for $f_s < 0$ saturation occurs only if negative forces (which assist motion) are applied. This is discussed in the conclusion.

In summary, mean-field predictions are consistent with numerical simulations of the full model for the case of homopolymer DNA. An important result of our analysis is that the velocity depends weakly on how the external force is distributed in f_C and f_F.

SEQUENCE SENSITIVITY OF MOTION

We have verified that the generalized Markov model has resolved the difficulties with the simple polymerization model. In this section we will apply our model to the study of sequence sensitivity. Recall that it is known that signals for pausing, arrest, or termination of transcription are encoded in the DNA sequence. This requires the RNAP molecule to be sensitive to extended base-pair patterns and not just to the nature of the one or two base pairs near the catalytic site (Chamberlin, 1995; Nudler et al., 1995). In the next several paragraphs we will first show that such an extended sequence sensitivity can be introduced in a natural way into our model, namely through the clamp-controlled activation barrier for sliding. In the subsequent sections we will perform model optical trap and footprinting simulations to test the model.

Modeling sequence sensitivity

The molecular mechanisms that are responsible for sequence sensitivity of RNAP are not known, but it is reasonable to assume that there must be a sequence-dependent noncovalent binding of groups of DNA or RNA bases to binding sites on the enzyme. This assumption implies that the binding energies as well as energy barriers that separate

discrete positions of binding sites must depend on the type of bases involved in the interactions. This provides a natural mechanism for the polymerase to "read" sequence patterns.

The observation of sequence sensitivity of elongation in principle does not tell us by itself whether the sequence information is "read" from the DNA template or from the RNA strand being polymerized. Consider, for example, the model shown in Fig. 1, *a* and *b*. Binding sites for DNA exist in the front region *F*, and a tight binding site *T* for RNA is involved in the process of threading out of the polymerized chain. As elastic stress is relaxed (while the position of the catalytic site is fixed), both the front *F* and the binding site *T* move with respect to the DNA and RNA base sequences, respectively. During this process both binding sites could be involved in detecting sequence patterns. Whereas bases "read" by the front are expected to be located downstream with respect to the catalytic site, bases read by the RNA tight binding site are located upstream (they have already been polymerized).

Using our generalized Markov model introduced in the third section, we can introduce sequence sensitivity by allowing the transition rates $k^{\pm}$ and $r^{\pm}$ to depend on the position n along the sequence $k^{\pm}(m, n)$ and $r^{\pm}(m, n)$. Polymerization step I takes place in the vicinity of the catalytic site, and we therefore assume that it is sensitive only to the bases that are in its vicinity (i.e., the base being polymerized and only a few neighbors). The stress relaxation process II, on the other hand, depends on the binding of RNAP to a larger sequence of bases within a strong binding site. This effect is described in our model by the value Λ of the energy barrier that has to be overcome to move RNAP in the presence of this binding. From this observation, we conclude that 1) the sequence sensitivity of process I is unlikely to allow for recognition of larger sequence patterns (5–10 bp), and 2) the sequence sensitivity of process II permits efficient extended sequence pattern recognition. This follows from the fact that as a larger number of strongly bound bases contribute, the corresponding variations of the energy barrier Λ can be larger. It is true that different bases that are polymerized in general have slightly different kinetic coefficients in process I, but corresponding energy differences are on the order of $\sim k_B T$ and should therefore not allow for significant velocity changes.

We assume in the following that sequence-dependent effects on process II dominate and that the sequence sensitivity of process I can be neglected. To define the dependence of the rates $r^{\pm}$ of process II on sequence, we denote the number of bases strongly bound to the polymerase by n_c. A sequence of length n_c therefore does influence motion and is "read" by the enzyme. We choose to locate the sites of binding to this sequence near the front *F* of the enzyme, where the double helix is opened. In principle, one could locate the selected bases also at other positions within the molecule, and, e.g., consider the case where the sequence of the newly polymerized RNA is detected at the RNA tight binding site *T*.

Motion of the front *F* by one step requires 1) the n_c base pairs unbind, 2) the front moves up 1 bp, and 3) rebinding to the new set of n_c base pairs. The energy activation barrier for this process is the parameter Λ introduced in the previous section. In general, this barrier depends on the precise sequence of bases bound in the n_c binding sites. To investigate the qualitative effects of sequence dependence, we study a simplified situation in which Λ is the sum of individual barriers for the four different types of bases (we denote these barriers by Λ_A, Λ_T, Λ_G, and Λ_C for A, T, C, and G bases). This assumption corresponds to the case where all binding sites are part of one rigid structure and thus unbind and rebind simultaneously. Flexibility of the enzyme could allow for a more complex process of stepping, where binding sites can unbind and rebind subsequently. In such a case, simply summing individual energy barriers would overestimate the correct value of Λ. However, Λ can still be much larger than $k_B T$ if many bases are involved, an important motivation for the choice of our simplified model.

This leads to the following sequence-dependent values for the energy barrier involved in one stepping event:

$$\Lambda(m, n) = \sum_{i=0}^{n_c-1} \Lambda_{B(n+1-m-i).} \tag{34}$$

Here $B(n)$ denotes the base at the position n along the DNA, and $n + l - m$ is the position of the front. The sequence-dependent energy barrier then reads

$$\Delta U_{II}(m, n) = \Lambda(m, n) - \frac{1}{2}(\mathscr{E}(m + 1) - \mathscr{E}(m)) + \frac{1}{2} a f_F. \tag{35}$$

Another possible source of sequence sensitivity is the free energy difference $\Delta G_{II}(m, n)$. However, variations in ΔG_{II} resulting from the DNA sequence are on the order of the binding energy of a single base, which we estimate to be on the order of $k_B T$. Thus, in contrast to the value of the barrier Λ, which can be large if many base pairs contribute, the effect of DNA sequence on ΔG_{II} is small. Sequence recognition is thus assumed in our model to be entirely incorporated into the sequence dependence of the activation barrier against sliding. An important and testable consequence is that the RNAP molecule should have no sequence recognition capacities under conditions of thermal equilibrium at the thermodynamic stall force $f = f_{stall}$. The reason is that reaction activation barriers do not affect the thermodynamic equilibrium state of a system. Recall also that we found earlier that for retarding forces that exceed f_s, the velocity is exponentially sensitive to the value of Λ. This provides a quite natural mechanism for influence by the base pair sequence on the polymerase dynamics.

To demonstrate this effect, we performed simulations of our model, moving along a random DNA sequence. For simplicity, we use a sequence with two different bases, A and G, only. As an example, we used $n_c = 6$, $\Lambda_G = 3/2 k_B T$,

$\Lambda_A = 1/3k_BT$, and $\Delta\mu = 10k_BT$. In this case, there exists a maximum barrier $\Lambda = 9k_BT$ if six or more G's occur subsequently. At this point, motion is very slow. Assuming a completely random sequence, the density of such patterns is $1/2^6 \simeq 10^{-2}$. For "typical" sequence patterns, elongation proceeds on the other hand rather fast. An example of a simulation of this case is shown in Fig. 5. Near positions $n = 100$ and $n = 200$, motion is slowed as a result of sequence patterns containing several G bases.

Fig. 6 shows an example of motion along a random sequence of bases A and G. The parameters used are $\alpha = 0.4$, $l_0 = 25$, $\Lambda_A = 1/3k_BT$, $\Lambda_G = 2/3k_BT$, $\Delta G = 10k_BT$, and $k_0/r_0 = 2$. The first and last lines represent the sequence chosen. The other lines indicate the position of the enzyme obtained for different times. The position of the catalytic site is denoted by *C*, the position of the front by *F*. Looking at the dynamics of model A shown in Fig. 6 *A*, we note that the catalytic site fluctuates forward and backward, whereas the front only moves forward. This is different for model B, in which the additional energy barrier disfavors depolymerization fluctuations (see Fig. 6 *B*). The molecule is less compressed for model B ($\langle m \rangle \simeq 7$) compared to model A ($\langle m \rangle \simeq 12$). In both cases, a slowing down is observable as *F* passes a pattern of several G-sites. However, from our simulations we observe that within our model the typical molecular compression does not fluctuate strongly along the sequence. In particular, our model does not show inchworm-like motion, as proposed by Chamberlin (1995) as a possible interpretation of footprinting experiments.

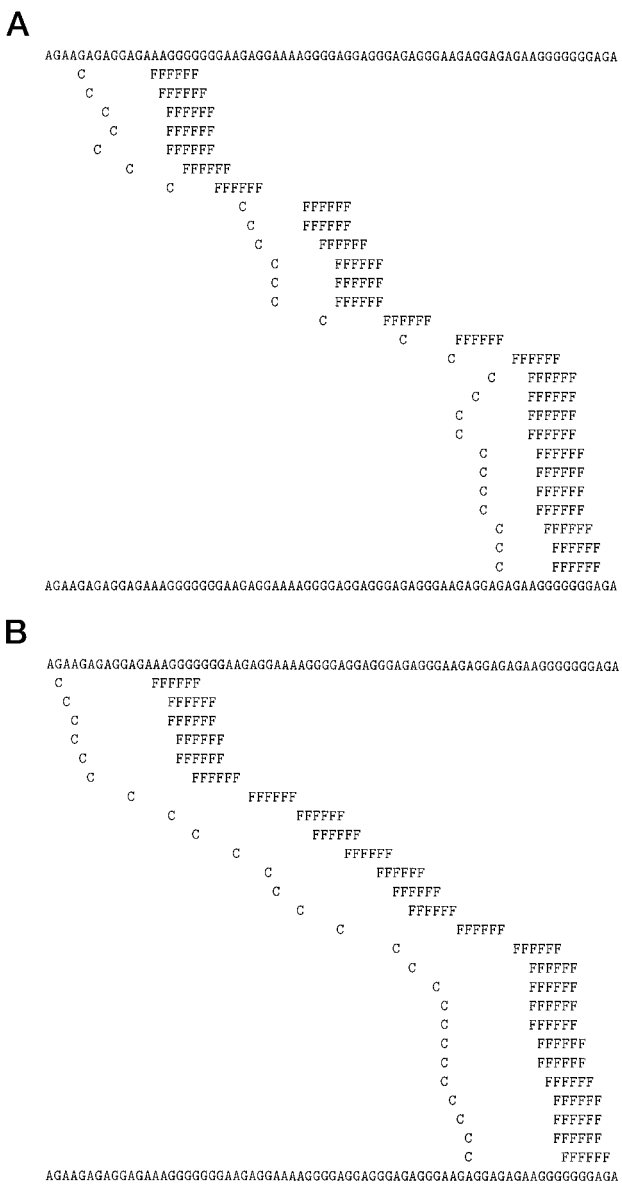

FIGURE 6 (*A*) Example of motion along a random sequence of two bases G and A, using the dynamics of model A. The first and last lines represent the sequence. The other lines indicate the positions *C* of the catalytic site and *F* of the clamp section at the front of the enzyme for different times. The time interval Δt between two subsequent lines corresponds to $\Delta t r_0 = 25$. (*B*) Same simulation, but for the dynamics of model B. In this case motion is slower and $\Delta t r_0 = 250$.

Stress relaxation and footprinting studies

Footprinting techniques allow us to detect those bases that are protected by the presence of the enzyme during elongation. Such a footprint of RNAP on DNA is obtained for chemically halted complexes, which are prepared by removing one type of NTP from the solution. As the corresponding complementary base along the DNA is reached, the enzyme forms a stalled complex from which a footprint can be obtained. Using the stochastic model described in the previous sections, we can study the stress relaxation that occurs in chemically halted complexes by mimicking footprinting studies. Even though the model neglects the possibility of detachment of the catalytic site we discussed in the Introduction, it allows us to qualitatively study some of the stress relaxation phenomena that are expected to be important in footprinting studies. We mimicked the footprinting technique within our model by first simulating the preparation of stalled complexes. This is done by defining a blocked site at position $n = n_b$, for which the forward rate vanishes ($r^+(m, n_b) = 0$). As the blocked site is reached, the system relaxes internal stresses during a time t_r, which

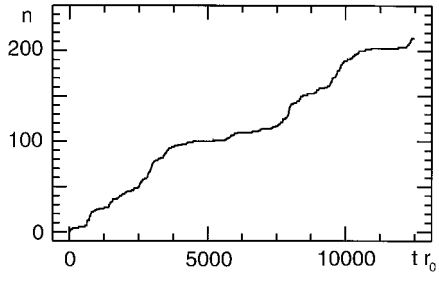

FIGURE 5 Position *n* as a function of time *t* along a random sequence of bases A and G for energy barriers $\Lambda_A = 1/3k_BT$, $\Lambda_G = 2/3k_BT$, chemical driving force $\Delta\mu = 10k_BT$, and $k_0/r_0 = 2$. Sequence sensitivity leads to pronounced fluctuations in velocity.

corresponds to the time during which stalled complexes are prepared. After relaxation, the analog of a footprint is taken during a second observation time t_f. In our model, we assume for simplicity that this "footprint analog" corresponds to the sites between the front F and the catalytic site C, which is motivated by the idea that changes in footprint size are related to elastic deformations of the enzyme.

Fig. 7 shows examples of footprint simulations using a random sequence of two bases A and G for both models A and B. The random sequence is displayed several times for different positions of the blocked site, which is indicated by an asterisk. The extremal positions of front and catalytic site obtained by the procedure described above are indicated for each case by C and F. As before in Fig. 6, the front is represented by several symbols indicating those sites that contribute to sequence sensitivity. The simulations were performed for the same parameters as in Fig. 6, with $t_r r_0 = 125$ and $t_f r_0 = 250$ for the example shown in Fig. 7 *A*, and $t_r r_0 = 250$, $t_f r_0 = 250$ for that in Fig. 7 *B*.

The figures show that these model footprints show clear variations that are due to sequence sensitivity. This effect is more pronounced for model A than for model B, which follows from the fact that for model A the initial stresses are larger on average than in the case of model B. Importantly, a comparison between Figs. 6 and 7 demonstrates that actual dynamical intermediates or "snapshots" of motion (as shown in Fig. 6) are very different from the footprint analogs shown in Fig. 7.

These observations show that during motion the velocity is strongly sequence dependent, but that the internal stress does not change very much. According to our model, after chemical stalling of the complex, this internal stress relaxes in a sequence-dependent manner, which produces the observed large variation in footprint sizes. The amount of stress relaxation depends on the available time. For very long times, $t_r \to \infty$, all stresses are relaxed and the final conformation is independent of DNA sequence. The reason is that under conditions of thermal equilibrium, sequence recognition is not possible for our model, as noted earlier. If, however, internal energy barriers Λ are sufficiently large compared to $k_B T$, then typical observation times may not be sufficient to complete stress relaxation. As a result, "frozen" intermediates of the relaxation process are observed that show a sequence-dependent size.

This argument is very general in that it is independent of the actual conformational changes that lead to the observed variations in footprint sizes. Even though we do not know the detailed microscopic origins of footprint sizes, our study suggests that stress relaxation phenomena are important for understanding footprinting experiments. The qualitative similarity between biochemical footprint studies and our simplified "model footprints" is, in fact, striking (for a comparison, see, e.g., Chamberlin, 1995). In summary, we can conclude that there is actually no contradiction between the generalized Markov model and footprint observations.

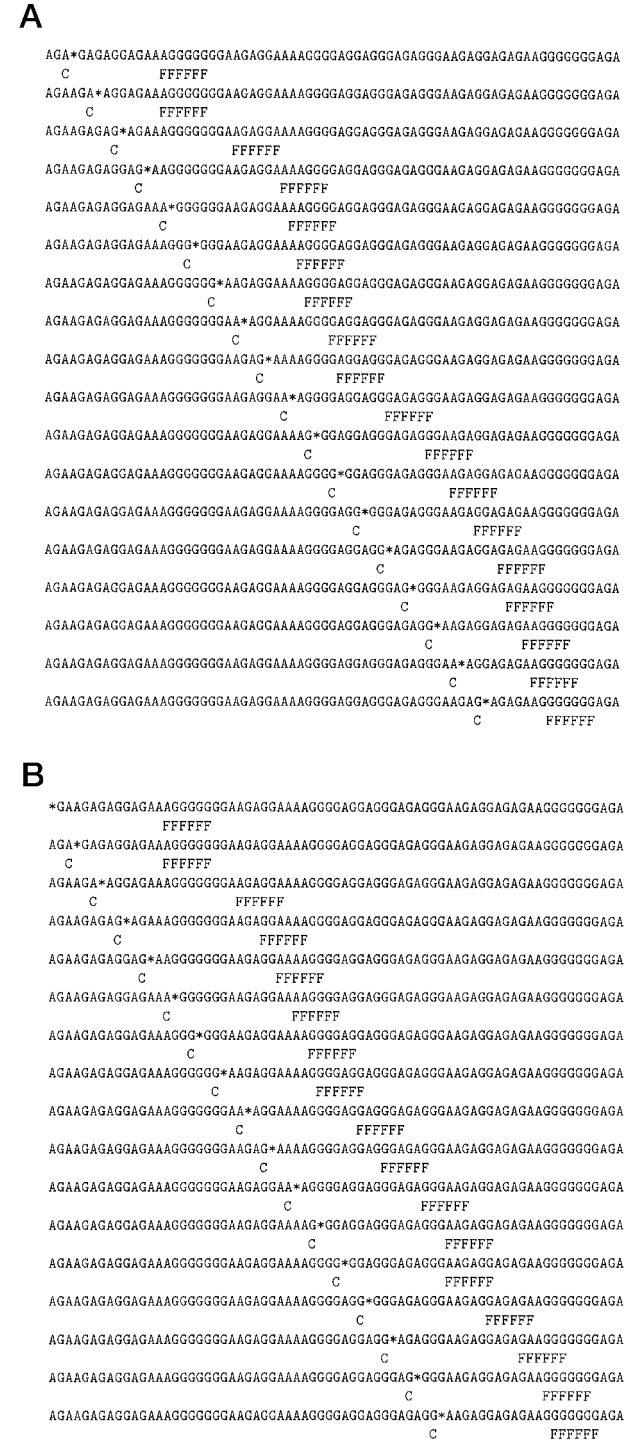

FIGURE 7 (*A*) Simulated "footprint analog," using the dynamics of model A. The base sequence is random and contains two bases A and G only. Along the base sequence, motion is halted at a blocked site (*marked by an asterisk*). The positions of the catalytic site C and the front F after a finite relaxation time are indicated. For details, see text. (*B*) Same simulation, but using the dynamics of model B.

Stalling experiments

Using our model, we can also study the analog of mechanical stalling experiments where the polymerase works

against an elastic element (represented by the optical trap in the experiment), thus building up a counter-force until motion stops. We model the trap by a spring that creates a force $f = -K(n - n_0)a$ opposing motion. Here n_0 is the position in base pairs along the DNA for which this spring is relaxed, n is the actual position, and K is the elastic modulus that corresponds to the optical trap. An example of such a stalling situation along an inhomogeneous segment of DNA is displayed in Fig. 8 for $n_0 = 0$ and $Ka^2/k_BT = 0.05$, with model A dynamics otherwise using the same parameter values as before. The external force is chosen to act only on the catalytic site (i.e., $f = f_C, f_F = 0$). For this choice, thermodynamic stalling conditions occur for $n = 200$. This example shows the typical behavior: a plateau where motion seems to stop for $n \simeq 100$ below stalling conditions. After a while, motion resumes a bit and a second plateau is reached. After a sufficiently long time, eventually the maximum elongation $n = 200$ is reached for thermodynamic stalling conditions.

These observations demonstrate the relevance of the force f_s as an apparent stalling force for practical conditions. As introduced earlier, for forces obeying $f_s < f < f_{stall}$, only a slow creeping motion remains, which for practical purposes may be indistinguishable from complete stalling. The force f_s for which this apparent stalling occurs is, according to Eq. 32, linearly dependent on the energy barrier for stress relaxation Λ and therefore strongly sequence dependent. For a given position along the DNA there exists a value of the stalling force characteristic of the local sequence. The apparent stall force thus must be regarded as a statistical quantity. Thus repeated stalling experiments will lead to a distribution of the observed stall forces that reflect the distribution of f_s along the sequence.

The above argument is valid for the case where energy barriers are large. The time available for observation is an important parameter (just as for the chemical stalling experiments of the previous section). If we perform a mechanical stalling experiment over a long time, creep motion can slowly lead to a buildup of the generated force until the force f_{stall} is reached. This stalling condition is not sequence dependent, and we therefore conclude that the fluctuations of observed stalling forces should decrease if very long observation times are available.

To demonstrate this effect, we performed repeated stalling simulations using an elastic spring attached to a randomly chosen site n_0. For each run a stall force was obtained by first allowing the enzyme to move over a time t_o until it seemed to have stalled. Afterward, the average force of the elastic element is measured during a second period of time t_o. The observed stalling forces can be represented in a histogram (see Fig. 9). The two histograms shown in Fig. 9, *A* and *B*, demonstrate how observed stalling forces depend on the observation time t_o. For sufficiently short observation times, there is a broad distribution of observed forces that are below the thermodynamic stalling force f_{stall} (see Fig. 9 *A*). As observation times are increased, the distribution peaks narrow at the force f_{stall}, as expected (see Fig. 9 *B*).

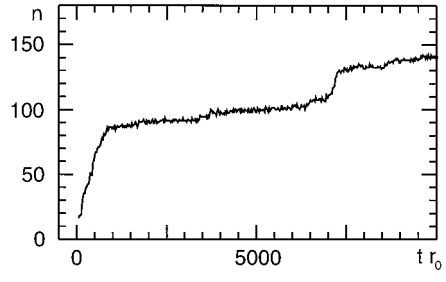

FIGURE 8 Position n as a function of time t as an example for a simulated stalling experiment in a trap. The system moves along a random base sequence while extending an elastic element. Thermodynamic stalling conditions are expected for $n = 200$. As a result of the base sequence, motion is irregular, showing pronounced plateaus where slow creep motion occurs before complete thermodynamic stalling is reached.

DISCUSSION

In this paper we propose a simple model for the elongation of RNA polymerase that can be used to analyze both chemical and mechanical stalling experiments. The central assumptions were as follows:

1. Mechanical force is generated by chemical energy transduction through the classical RNA polymerization reaction.
2. Both the polymerization reaction and the RNAP sliding motion are inhibited by internal strain or tension of the molecule.
3. Sequence recognition of extended base pair patterns is based on sequence sensitivity of the RNAP energy barriers against sliding.

In addition, we had to make assumptions on the nature of the force-transduction mechanism. We studied two limiting cases, direct energy transduction (model A) and the case of strain-dependent energy barriers (model B).

Each of the assumptions 1–3 is reasonable in the framework of our current understanding of RNAP, but they certainly can be questioned. Experimental tests of 1 and 3 appear to be possible. To test 1, one could, for instance, study whether fully stalled RNAP complexes still consume NTPs through fluorescence labeling experiments. Our model predicts that this should not be the case. Assumption 3 could be tested by measuring the distribution of (apparent) stalling forces as a function of the observation time. Our theory predicts that this distribution should narrow for longer observation times. However, there does not appear to be an obvious way to test assumption 2 directly.

The central prediction of our model is that the stall force measured in the mechanical stalling experiments can be significantly smaller than the thermodynamic stall force, and that unlike the thermodynamic stall force, the apparent stall force is both sequence dependent and dependent on the observation period. This could be tested by performing stalling experiments on DNA sequences with varying dis-

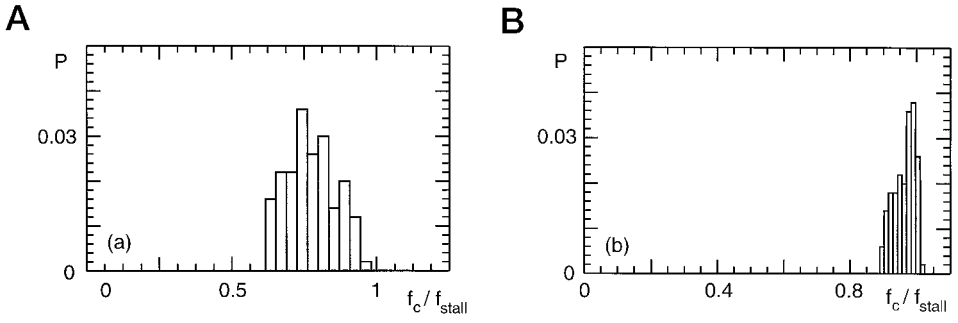

FIGURE 9 (*A*) Histogram of apparent stalling forces f_C measured in units of the thermodynamic stalling force f_{stall}. Apparent stalling forces were obtained by observing 100 stalling events after an observation time $t_o r_0 = 2500$ and averaging the generated force during a second interval t_o. (*B*) Same histogram, but for $t_o r_0 = 25{,}000$.

tributions of base pairs. If stalling is thermodynamic in nature, then the stalling force should be dependent only on the base pairs at or close to the catalytic site *C*. The distribution of measured stall forces would then depend only weakly on the nature of the base pair sequence. If, on the other hand, the observed stalling is only apparent, as proposed by our model, then the sequence dependence of stall forces should be pronounced and the stall force distribution should narrow with increasing observation time.

There predictions were of a qualitative nature. Quantitative tests of the proposed model could be performed using homopolymeric DNA. For this case we find that the average velocity depends only weakly on how the total applied force f is divided in contributions f_C and f_F acting at different locations within the enzyme. This is important because the stress distribution within the enzyme should depend on the details of attachment of RNAP to a substrate, which are difficult to control under experimental conditions. Our result therefore suggests that meaningful force-velocity curves of RNAP can be obtained even if details of molecular attachment to the substrate are unknown. Furthermore, we predict that for high force levels ($f \gg f_s$), the average velocity v becomes exponentially small: $v \simeq v_0 \exp(-af/2k_BT)$, where v_0 is independent of f. Careful studies of RNAP "creep" should be an unambiguous quantitative test of the model.

Another interesting area in which tests are possible concerns the fluctuations of the RNAP position $\langle n \rangle$ around its mean position, or alternatively, measurement of the effective diffusion constant D. This quantity can in principle be measured by fluctuation analysis of the optical trap experiment. According to the simple polymerization model of the second section, we expect (for no external load) that $D \simeq va/2$. For heteropolymer DNA the situation is more complex. The motion of RNAP along an irregular base pair sequence is formally similar to the motion of a particle in a linear random potential, which is a model for a disordered environment (Alexander et al., 1981; Bernasconi and Schneider, 1983; Marinani et al., 1983). In this case, fluctuations are enhanced by the disorder. Sequence-induced velocity fluctuations in our model play a similar role. Our results show that sequence effects can lead to a dramatic increase in velocity fluctuations, which increases the effective diffusion coefficient. In principle, nondiffusive behavior of the variance of motion could also be possible as a result of structured sequences.

As an important general question, it could well be asked whether simple chemical kinetics models such as the one proposed in this paper could really be hoped to describe a very large and complex protein like RNAP. The binding of RNAP to the promoter site is indeed well known to be under extensive biochemical control. We are encouraged, however, by the fact that RNAP molecules from different sources can elongate along different DNA sequences, which suggests that the basic copying "machinery" may be less complex. As mentioned, descriptions similar to the one presented in this paper have also proved useful for describing DNA polymerase (Creighton and Goodman, 1995).

If simple models such as the one proposed in this paper are found experimentally to be able to account for the essential properties of RNAP motion, then this would provide important evidence for a relatively simple and universal underlying mechanism for the elongation stage of transcription.

APPENDIX A: MEAN-FIELD THEORY

To find solutions to the mean-field Eq. 29, we introduce the variables

$$\begin{aligned} \delta &\equiv \exp(-[\Lambda + af_F/2]/k_BT) \\ \epsilon &\equiv \exp(-[\Delta\mu - af_C]/k_BT) \\ \gamma &\equiv \exp([\Delta E + af_F]/k_BT) \\ \rho &\equiv \exp(-f_C a/k_BT). \end{aligned} \tag{A1}$$

For model A (direct transduction), Eq. 29 can then be written as

$$\bar{k}_0(1 - \epsilon x^2) = r_0 \delta\, x(1 - \gamma/x^2) \quad \text{for} \quad \Delta U_{II} > 0 \tag{A2a}$$

$$\bar{k}_0(1 - \epsilon x^2) = r_0(1 - \gamma/x^2) \quad \text{for} \quad \Delta U_{II} < 0, \tag{A2b}$$

(Model A)

and for model B we obtain

$$\bar{k}_0\rho(1/x^2 - \epsilon) = r_0 \delta\, x(1 - \gamma/x^2) \quad \text{for} \quad \Delta U_{II} > 0 \tag{A3a}$$

$$\bar{k}_0\rho(1/x^2 - \epsilon) = r_0(1 - \gamma/x^2) \quad \text{for} \quad \Delta U_{II} < 0, \tag{A3b}$$

(Model B)

with $\bar{k}_0 \equiv k_0[\mathrm{NTP}]$. The unknown variable x in Eqs. A2 and A3 is related to the mean-field value of $\langle m\rangle$ (recall that $\mathcal{E}(m) = \alpha m^2$) according to

$$\langle m\rangle + \frac{1}{2} = \frac{k_B T}{\alpha} \ln x. \tag{A4}$$

This value of $\langle m\rangle$ then allows us to self-consistently determine the sign of ΔU_{II}. The mean-field velocity $v = d\langle n\rangle/dt$ can be calculated according to Eq. 30. Using the solution for x to Eqs. A2 and A3, this leads to

$$v = \bar{k}_0 a(1 - \epsilon x^2) = r_0 \delta x(1 - \gamma/x^2) \quad \text{for} \quad \Delta U_{II} > 0 \tag{A5a}$$

$$v = \bar{k}_0 a(1 - \epsilon x^2) = r_0(1 - \gamma/x^2) \quad \text{for} \quad \Delta U_{II} < 0 \tag{A5b}$$

(Model A)

and

$$v = \bar{k}_0 a\rho(1/x^2 - \epsilon) = r_0 \delta x(1 - \gamma/x^2) \quad \text{for} \quad \Delta U_{II} > 0 \tag{A6a}$$

$$v = \bar{k}_0 a\rho(1/x^2 - \epsilon) = r_0(1 - \gamma/x^2) \quad \text{for} \quad \Delta U_{II} < 0. \tag{A6b}$$

(Model B)

From Eqs. A2 and A3 we observe that stalling conditions $v = 0$ imply that $\gamma = 1/\epsilon$, which leads for both models A and B to the thermodynamic stalling force given by Eq. 31. In writing Eqs. A2 and A3, we have implicitly assumed that ΔG_I is positive for $m = \langle m\rangle$. This condition implies (see Eq. 19) that we restrict ourselves to complexes that move forward, i.e., $f \le f_{stall}$.

General results for model A

The mean-field Eqs. A2 and A3 are polynomial expressions in x that can, in principle, be solved directly. We will not discuss the general solution, but restrict ourselves in the following to the particular case of model A with $\bar{k}_0 > r_0$, i.e., a polymerization rate that is faster than the maximum stress relaxation rate.

For this case we can identify interesting limiting cases and extract general results. First we look at the situation of small external forces.

Weak forces: $f \ll f_s$

Let $f = f_C + f_F$ be the total applied force. We will say that f is in the weak force regime if $f \ll f_s$, where $f_s = (\Delta\mu - 2\Lambda)/a$. Under practical conditions, the force f_s is small compared to the thermodynamic stall force (Eq. 31). It will be shown below that in the weak force regime $\Delta U_{II} < 0$, which implies that stress relaxation is not activated. Using Eq. A2b, we obtain for the quantity x (which is related to $\langle m\rangle$ by Eq. A4)

$$x = \left(\frac{\bar{k}_0 - r_0 + \sqrt{(\bar{k}_0 - r_0)^2 + 4\bar{k}_0 r_0 \epsilon\gamma}}{2\bar{k}_0\epsilon}\right)^{1/2}. \tag{A7}$$

In the weak force regime $\epsilon\gamma \ll 1$, and we find that $\langle m\rangle$ has a linear dependence on f_C:

$$\langle m\rangle + \frac{1}{2} \simeq \frac{1}{2\alpha}(\Delta\mu - af_C) + \frac{k_B T}{2\alpha} \ln(1 - r_0/\bar{k}_0). \tag{A8}$$

For the velocity, we obtain

$$v \simeq ar_0\left(1 - \frac{\bar{k}_0}{\bar{k}_0 - r_0} \exp(a[f - f_{stall}]/k_B T)\right), \tag{A9}$$

describing the saturation of velocity at the maximum value $v_{max} = ar_0$, assuming that $f_{stall}/k_B T \gg 1$.

Intermediate forces: $f_{stall} \gg f \gg f_s$

For $f \gg f_s$, ΔU_{II} becomes positive and stress relaxation is activated. We now have to use Eq. A2a. Here we restrict ourselves to the case $f \ll f_{stall}$, where $\Delta G_{II}/k_B T \ll 0$ and $\gamma \simeq 0$. This allows us to neglect the last term in Eq. A2a. In this regime, strain loading, i.e., increases in m, induced by the external force f_F plays no role. The only physical solution to Eq. A2a with $\gamma = 0$ leads to $x = (-\bar{\delta} + (\bar{\delta}^2 + 4\epsilon)^{1/2})/2\epsilon$ and

$$v \simeq \bar{k}_0 a\left(1 - \frac{1}{4\epsilon}[-\bar{\delta} + \sqrt{\bar{\delta}^2 + 4\epsilon}]^2\right), \tag{A10}$$

where $\bar{\delta} \equiv \delta r_0/\bar{k}_0$. For $f \gg f_s$, $\delta^2/4\epsilon \ll 1$, and we can expand in powers of $\bar{\delta}/2\epsilon^{1/2}$:

$$\begin{aligned}\langle m\rangle + \frac{1}{2} &\simeq -\frac{k_B T}{2\alpha}\left(\frac{\bar{\delta}}{\epsilon^{1/2}} + \ln\epsilon\right) \\ &= \frac{1}{2\alpha}(\Delta\mu - af_C) - \frac{k_B T}{2\alpha}\frac{r_0}{\bar{k}_0}\exp(a[f_s - f]/2k_B T).\end{aligned} \tag{A11}$$

Note that $\langle m\rangle$ still depends linearly on f_C. The velocity vanishes exponentially with f for $f \gg f_s$, as described by Eq. 33.

We still have to show that our assumptions that $\Delta G_I(\langle m\rangle) > 0$ for the mean-field value $\langle m\rangle$ and $\Delta G_{II}(\langle m\rangle) \ll -k_B T$ for the regime of intermediate forces are correct.

First we check the assumption for ΔG_{II}. In the weak force regime $f \ll f_s$, using Eq. A8,

$$\Delta G_{II}(\langle m\rangle) = \Delta E + af - \Delta\mu. \tag{A12}$$

Assuming that the energy ΔE is on the order of a hydrogen bond energy (energy cost to "unwind" DNA by 1 bp) and thus is on the order of $k_B T$. Assuming that $\Delta\mu \gg k_B T$, we find that $\Delta G_{II} \ll -k_B T$. In the limit of intermediate forces, we find, using Eq. A11 again,

$$\Delta G_{II}(\langle m\rangle) \simeq \Delta E + af - \Delta\mu. \tag{A13}$$

Thus the condition on $\Delta G_{II} \ll -k_B T$ holds as long as $f \ll f_{stall}$. For $f \simeq f_s$, we find that $\Delta G_{II} \simeq \Delta E - 2\Lambda$. Using our estimate of $\Lambda = 4k_B T$, we conclude that our approximation is valid for forces around $f \simeq f_s$.

Another assumption is that ΔG_I is positive. In the weak force regime, we find that

$$\Delta G_I(\langle m\rangle) \simeq k_B T - \ln(1 - r_0/\bar{k}_0), \tag{A14}$$

which is positive for $\bar{k}_0 > r_0$. In the intermediate force limit,

$$\Delta G_I \simeq k_B T \exp(a[f_s - f]/k_B T) \tag{A15}$$

is positive albeit small.

Finally, we have to check the sign of ΔU_{II}. In the regime of weak forces, $f \ll f_s$, using Eqs. 25 and A8, ΔU_{II} is negative:

$$\Delta U_{II}(\langle m\rangle) \simeq \Lambda - \Delta\mu/2 + f/2. \tag{A16}$$

For $f \simeq f_s$, ΔU_{II} becomes positive in the regime of intermediate and large forces.

APPENDIX B: MONTE CARLO SIMULATION

The model described in the third section is defined in terms of transition rates $k^{\pm}$ and $r^{\pm}$, which give probabilities per unit time for the occurrence of steps corresponding to processes I and II. Assuming that at time t the system is in the state (n, m), there exist finite probabilities $P^{I}_{\pm}$ and $P^{II}_{\pm}$ that

at time $t + \Delta t$ transitions corresponding to processes I or II in the forward or backward direction have occurred. For small Δt we can write

$$P^{\mathrm{I}}_{+} \simeq \Delta t\, k^{+}(m) \quad \text{for} \quad (n, m) \rightarrow (n+1, m+1) \tag{B1}$$

$$P^{\mathrm{I}}_{-} \simeq \Delta t\, k^{-}(m) \quad \text{for} \quad (n, m) \rightarrow (n-1, m-1) \tag{B2}$$

$$P^{\mathrm{II}}_{+} \simeq \Delta t\, r^{+}(m) \quad \text{for} \quad (n, m) \rightarrow (n, m+1) \tag{B3}$$

$$P^{\mathrm{II}}_{-} \simeq \Delta t\, r^{-}(m) \quad \text{for} \quad (n, m) \rightarrow (n, m-1), \tag{B4}$$

where terms higher order in Δt have been ignored. The probability that no event will occur is $P_0 \simeq 1 - \Delta t\,(k^{+}(m) + k^{-}(m) + r^{+}(m) + r^{-}(m))$. Equations B1–B4 correspond to a discretized version of the original stochastic process. It can be studied numerically by deciding according to the probabilities $P^{\pm}_{\mathrm{I}}$ and $P^{\pm}_{\mathrm{II}}$ if one of the four possible steps occurs. For each step, a random number $0 \le r < 1$ is drawn. Depending on its value, a decision is taken as follows:

$$(n, m) \rightarrow (n+1, m+1) \quad \text{if} \quad 0 \le r < P^{\mathrm{I}}_{+} \tag{B5}$$

$$(n, m) \rightarrow (n-1, m-1) \quad \text{if} \quad P^{\mathrm{I}}_{+} \le r < P^{\mathrm{I}}_{+} + P^{\mathrm{I}}_{-} \tag{B6}$$

$$(n, m) \rightarrow (n, m+1) \quad \text{if} \quad P^{\mathrm{I}}_{+} + P^{\mathrm{I}}_{-} \le r < P^{\mathrm{I}}_{+} + P^{\mathrm{I}}_{-} + P^{\mathrm{II}}_{+} \tag{B7}$$

$$(n, m) \rightarrow (n, m-1) \quad \text{if} \quad P^{\mathrm{I}}_{+} + P^{\mathrm{I}}_{-} + P^{\mathrm{II}}_{+} \le r < P^{\mathrm{I}}_{+} + P^{\mathrm{I}}_{-} + P^{\mathrm{II}}_{+} + P^{\mathrm{II}}_{-}. \tag{B8}$$

As a result, the values $n(t_{\mathrm{i}})$ and $m(t_{\mathrm{i}})$ are obtained. Here $t_{\mathrm{i}} = i\Delta t$ are the discrete times. Small time steps are used to choose a sufficiently small acceptance rate of transitions, $a_{\mathrm{r}} = (P^{\mathrm{I}}_{+} + P^{\mathrm{I}}_{-} + P^{\mathrm{II}}_{+} + P^{\mathrm{II}}_{-})\Delta t \ll 1$. For long runs ($i = 1{:}N$, N large), we measure the average velocity,

$$v \equiv \frac{a}{N}\,(n(t_{\mathrm{N}}) - n(t_1)), \tag{B9}$$

the average compression,

$$\langle m \rangle \equiv \frac{1}{N}\sum_{\mathrm{i}=1}^{\mathrm{N}} m(t_{\mathrm{i}}), \tag{B10}$$

the diffusion coefficient,

$$D \equiv \frac{a^2}{N}\sum_{\mathrm{i}=1}^{N} \frac{(n(t_{\mathrm{i}}) - v t_{\mathrm{i}})^2}{t_{\mathrm{i}}}, \tag{B11}$$

and the compressional fluctuations,

$$\delta m \equiv \langle (m - \langle m \rangle)^2 \rangle^{1/2}. \tag{B12}$$

We thank S. Block, J. Gelles, P. Hansma, and M. Wang for helpful conversations.

RB thanks the Rothschild Foundation and the Department of Materials Research of the National Science Foundation (grant DMR-9407741) for financial support.

REFERENCES

Alexander, S., J. Bernasconi, W. R. Schneider, and R. Orbach. 1981. Excitation dynamics in random one-dimensional systems. *Rev. Mod. Phys.* 53:175–198.

Bernasconi, J., and W. R. Schneider. 1982. Diffusion in a one-dimensional lattice with random asymmetric transition rates. *J. Phys. A Math. Gen.* 15:L729–L734.

Chamberlin, M. J. 1995. New models for the mechanism of transcription elongation and its regulation. *Harvey Lect.* (Ser. 88):1–21.

Chan, C. L., and R. Landick. 1994. New perspectives on RNA chain elongation and termination by *E. coli* RNA polymerase. *In* Transcription: Mechanisms and Regulation. R. C. Conavay and J. W. Conaway, editors. Raven Press, New York. 297–321.

Creighton, S., and M. F. Goodman. 1995. Gel kinetic analysis of DNA polymerase fidelity in the presence of proofreading using bacteriophage T4 DNA polymerase. *J. Biol. Chem.* 270:4759–4774.

Derényi, I., and T. Vicsek. 1996. The kinesin walk: a dynamic model with elastically coupled heads. *Proc. Natl. Acad. Sci. USA.* 93:6775–6779.

Duke, T., and S. Leibler. 1996. Motor protein mechanics: a stochastic model with minimal mechanochemical coupling. *Biophys. J.* 71: 1235–1247.

Erie, D. A., T. D. Yager, and P. H. von Hippel. 1992. The single-nucleotide addition cycle in transcription: a biophysical and biochemical perspective. *Annu. Rev. Biophys. Biomol. Struct.* 21:379–415.

Hua, W., E. C. Young, M. L. Fleming, and J. Gelles. 1997. Coupling of kinesin steps to ATP hydrolysis. *Nature.* 388:390–393.

Jülicher, F., A. Ajdari, and J. Prost. 1997. Modelling molecular motors. *Rev. Mod. Phys.* 69:1269–1282.

Komissarova, N., and M. Kashlev. 1997. Transcriptional arrest: *Escherichia coli* RNA polymerase translocates backward, leaving the 3′ end of the RNA intact and extruded. *Proc. Natl. Acad. Sci. USA.* 94:1755–1760.

Krummel, B., and M. J. Chamberlin. 1992. Structural analysis of ternary complexes of *Escherichia coli* RNA polymerase. Desoxyribonuclease I footprinting of individual complexes along different transcription units. *J. Mol. Biol.* 225:239–250.

Landick, R. 1997. RNA polymerase slides home: pause and termination site recognition. *Cell.* 88:741–744.

Leibler, S., and D. Huse. 1993. Porters versus rowers: a unified stochastic model. *J. Cell Biol.* 121:1357–1368.

Lewis, B., Genes V. Oxford University Press, Oxford, 1994.

Marinani, E., G. Parisi, D. Ruelle, and P. Windey. 1983. Random walk in a random environment and 1/*f* noise. *Phys. Rev. Lett.* 50:1223–1225.

Metzger, W. M., P. Schicktor, and H. Heumann. 1989. A cinematographic view of *Escherichia coli* RNA polymerase translocation. *EMBO J.* 8:2745–2754.

Meyhöfer, E., and J. Howard. 1995. The force generated by a single kinesin molecule against an elastic load. *Proc. Natl. Acad. Sci. USA.* 92: 574–578.

Monod, J., J. Wyman, and J. P. Changeux. 1965. On the nature of allosteric transitions: a plausible model. *J. Mol. Biol.* 12:88–118.

Mustaev, A., M. Kashlev, E. Zaychikov, M. Grachev, and A. Goldfarb. 1993. Active center rearrangement in RNA polymerase initiation complex. *J. Biol. Chem.* 268:19185–19187.

Nudler, E., A. Goldfarb, and M. Kashlev. 1994. Discontinuous mechanism of transcription elongation. *Science.* 265:793–796.

Nudler, E., M. Kashlev, V. Nikiforov, and A. Goldfarb. 1995. Coupling between transcription elongation and RNA polymerase inchworming. *Cell.* 81:351–357.

Oosawa, F., and S. Asakura. 1975. Thermodynamics of the Polymerization of Proteins. Academic Press, London, New York.

Peskin, C. S., G. B. Ermentrout, and G. F. Oster. 1994. The correlation ratche: a novel mechanism for generating directed motion by ATP hydrolysis. *In* Cell Mechanics and Cellular Engineering, V. C. Mow, F. Guilak, R. Tran-Son-Tay and R. Hochmuth, editors. Springer Verlag, New York. 479–489.

Peskin, C. S., and G. F. Oster. 1995. Coordinated hydrolysis explains the mechanical behavior of kinesin. *Biophys. J.* 68:202–210.

Prost, J., J. F. Chauwin, L. Peliti, and A. Ajdari. 1994. Asymmetric pumping of particles. *Phys. Rev. Lett.* 72:2652–2655.

Reeder, T. C., and D. K. Hawley. 1996. Promotor proximal sequences modulate RNA polymerase II elongation by a novel mechanism. *Cell.* 87:767–777.

Rice, G. A., C. M. Kane, and M. J. Chamberlin. 1991. Footprinting analysis of mammalian RNA polymerase II along its transcript: an alternative view of transcript elongation. *Proc. Natl. Acad. Sci. USA.* 88:4245–4249.

Schnitzer, M., and S. Block. 1997. Kinesin hydrolyses one ATP per 8-nm step. *Nature.* 388:386–390.

Stryer, L. 1988. Biochemistry. W. H. Freeman, New York.

Svoboda, K., P. Mitra, and S. Block. 1994. Fluctuation analysis of motor protein movement and single enzyme kinetics. *Proc. Natl. Acad. Sci. USA.* 91:11782–11786.

Wang, D., T. Meier, C. Chan, G. Feng, D. Lee, and R. Landick. 1995. Discontinuous movements of DNA and RNA in RNA polymerase accompany formation of a paused transcription complex. *Cell.* 81: 341–350.

Yager, T. D., and P. H. von Hippel. 1987. Transcript elongation and termination in *Escherichia coli. In Escherichia coli* and *Salmonella typhimurium*. F. C. Neidhard, J. L. Ingraham, K. B. Low, B. Magasanik, M. Schaechter, and H. E. Umbarger, editors. 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.