Fluctuations, Pauses, and Backtracking in DNA Transcription

Margaritis Voliotis,*† Netta Cohen,* Carmen Molina-París,† and Tanniemola B. Liverpool†‡
*School of Computing, †Department of Applied Mathematics, University of Leeds, Leeds, United Kingdom; and ‡Department of Mathematics, University of Bristol, Bristol, United Kingdom

ABSTRACT Transcription is a vital stage in the process of gene expression and a major contributor to fluctuations in gene expression levels for which it is typically modeled as a single-step process with Poisson statistics. However, recent single molecule experiments raise questions about the validity of such a simple single-step picture. We present a molecular multistep model of transcription elongation that demonstrates that transcription times are in general non-Poisson-distributed. In particular, we model transcriptional pauses due to backtracking of the RNA polymerase as a first passage process. By including such pauses, we obtain a broad, heavy-tailed distribution of transcription elongation times, which can be significantly longer than would be otherwise. When transcriptional pauses result in long transcription times, we demonstrate that this naturally leads to bursts of mRNA production and non-Poisson statistics of mRNA levels. These results suggest that transcriptional pauses may be a significant contributor to the variability in transcription rates with direct implications for noise in cellular processes as well as variability between cells.

INTRODUCTION

It has long been appreciated that noise and fluctuations play an important role in the cellular environment (1). Small numbers of molecules as well as the intrinsically stochastic nature of biochemical reactions mean that fluctuations must be taken into account to understand cellular function. More recently there has been renewed interest in genetic noise (see, e.g., (2–4)) and fluctuations at the molecular level, driven by new observational techniques which allow one to track levels of chemical species in bacterial and yeast cells (5–7). These experiments have allowed the identification of a number of different sources of fluctuations in the expression levels of a particular gene. Low numbers of macromolecules that participate in gene regulation and expression, as well as macroscopic fluctuations in the environment, are likely to affect the statistics of gene expression. In addition, the stochastic nature of the production and degradation of RNA transcription products introduces an important source of intrinsic genetic noise.

Within the central dogma of molecular biology, gene expression can be split into two distinct phases, transcription of DNA to mRNA and translation of mRNA into protein. However, the production (and degradation) of proteins and mRNA transcripts are themselves multistage processes. Transcription, in particular, can be crudely broken up into three main stages: initiation, elongation, and termination. During initiation, RNA polymerase (RNAP) binds to a promoter sequence on the DNA and opens the double helix, uncovering the template strand to be transcribed. The subsequent transcription of the first few (8–12) nucleotides leads to the formation of the transcription elongation complex (TEC) which consists of the RNAP, the DNA, and the nascent mRNA (8). The formation of the TEC signals the entrance into the elongation phase where, under normal conditions, the TEC slides along the DNA, extending the transcript one nucleotide at a time. Destabilization of the TEC (at specific sites or by certain factors) leads to the termination of the process and the release of the nascent mRNA (9).

In fact, the transcription process can exhibit biochemical fluctuations at each stage and cannot, in general, be described by the simple exponential (Poisson) birth and death Markov processes that are currently used to analyze experiments ((4,10) and references therein). This naturally leads one to ask under what conditions is the Poisson approximation valid (11). To answer this question, a more detailed analysis of the dynamics of transcription is required. Recent single molecule experiments (12,13) also provide a new window into the dynamics of transcription, offering a motivation as well as a solid basis for constructing more detailed mathematical models.

As demonstrated below, implicit in the Poisson approximation for the stochastic description of transcription is the assumption that the rate-limiting step is initiation, i.e., that the time taken for the polymerase to find the promoter sequence by random diffusion is longer than the total time for elongation. If so, fluctuations in the initiation step would be the major contributor to genetic noise due to transcription.

In general, the frequency of transcription initiation has a wide dynamical range in vivo (14), and in vitro studies have shown that initiation times can be as fast as a few seconds (15–17). Clearly then, rapid initiation times can be significantly shorter than the time needed for elongation, especially for long DNA templates or bacterial genes transcribed in operons. In such cases, modeling transcription as a Markovian process, obeying Poisson statistics, may be an inadequate approximation. In fact, transcription elongation

Submitted February 2, 2007, and accepted for publication August 7, 2007.

Address reprint requests to Tanniemola B. Liverpool, E-mail: t.liverpool@bristol.ac.uk.

Editor: Michael Edidin.

0006-3495/08/01/334/15 $2.00

doi: 10.1529/biophysj.107.105767

demonstrates features that suggest that it could play as significant a role in the overall rate of transcription and hence the regulation of gene expression (18).

Of particular interest are transcriptional pauses that disrupt the processive mRNA synthesis. Single-molecule techniques have made a more quantitative characterization of elongation pauses possible. Recent in vitro experimental studies with *Escherichia coli* RNAP have classified elongation pauses into long (>20 s) and short (1–6 s) pauses (19,20). It has also been suggested that elongation pauses can occur either in a sequence-dependent manner (21) or irrespective of the underlying sequence (19) and some pauses were linked with the reverse translocation of the RNAP (backtracking) (19,22). Backtracking may be caused by nucleotide misincorporation or a weak RNA-DNA hybrid (8,23) and can also be regulated by specific proteins (24). In general, backtracking can significantly increase the total elongation time, and in many cases is the precursor to transcriptional arrest (25).

In this article we point out that a single step Poissonian picture of transcription implies that the rate-limiting step (in transcription) is transcription initiation, i.e., the elongation process that follows is fast and straightforward. We present a molecular model of transcription elongation (26–29) with very different, heavy-tailed distributions of transcription times. Furthermore, we show that elongation can be sufficiently slow to be rate-limiting, providing the cell with ample targets for regulation. In particular, we highlight the very important role transcriptional pauses play in determining the distribution of total transcription times and therefore the statistics of the mRNA levels. Our results should have direct implications for the fluctuations observed in the levels of gene expression, which lead to noise in cellular processes and may play a role in generating variability between cells.

We study two classes of models both analytically, within a mean field approximation, and numerically, using stochastic simulations. First in a model of transcription without transcriptional pauses (Model A), we find that the transcription-elongation adds a typical delay that scales linearly with the transcript size. In this model, the contribution from fluctuations is small (especially for large transcript lengths) and leads to elongation times that are described by a Gaussian distribution. Second, we construct a model that incorporates backtracking pauses during the elongation phase (Model B). We develop a detailed model of backtracking pauses as a first-passage process and study the distribution of their duration considering two different scenarios: 1), pauses that end with the TEC sliding back into position (case 1); and 2), backtracking pauses that can also lead to transcriptional arrest (case 2). In addition, using stochastic simulations, we investigate the effect of backtracking pauses on the distribution of elongation times, as well as on the statistics of the mRNA production. We show that pauses can dominate the elongation process and lead to a heavy-tailed distribution of elongation, and hence transcription completion times. Finally, we use Model B to perform simulations of mRNA production, allowing multiple RNAP molecules to transcribe the same gene. We demonstrate that rare and long-lived pauses result in bursts of mRNA production, in agreement with experimentally observed transcriptional bursting (11,30,31).

TRANSCRIPTION ELONGATION COMPLEX

At a typical template position the RNAP covers a region of ~25 DNA basepairs (bp), of which the central part (12 bp) is melted, forming the transcription bubble (32). Within the bubble, a hybrid (8–9 bp) is formed between the nascent mRNA and the complementary DNA strand that contributes to the stability of the TEC (33). Elongation (polymerization) describes the addition of a nucleotide to the 3′ end of the transcript, which is catalyzed by the active site of the RNAP and hence conditional on the active site being locked in the appropriate position. In the simplest scenarios, polymerization of the nascent mRNA can be interrupted by the reverse process of pyrophosphorolysis (depolymerization), which leads to shortening of the mRNA transcript (8), or by pauses, due to translocation of the TEC (see below).

After a polymerization step has taken place the TEC is thought to occupy the pretranslocated state. From this position the TEC must translocate forward on the DNA template, to the posttranslocated state, so that the active site is in position to catalyze the next nucleotide addition. In general, the TEC is also capable of translocating backward on the template (backtracking) or even ahead of the target DNA nucleotide (hypertranslocation). During backtracking the TEC is moved upstream along the DNA template. This translocation causes the 3′ end of the nascent mRNA to dissociate from the DNA and exit the TEC through the secondary channel of the polymerase (34). Effectively, this rearward motion dissociates the active site from the 3′ end of the transcript, temporarily halting the elongation, until the TEC is in position once again. The posttranslocated, pretranslocated, and backtracked states are illustrated schematically in Fig. 1, *a–c*.

A simple mathematical model that captures the essence of polymerization, depolymerization, and backtracking can be described in terms of two discrete variables n and m. Variable n denotes the position of the last transcribed nucleotide, or equivalently, the size of the nascent mRNA, and ranges from 0 to N. In our model, n counts nucleotides relative to the position at which the elongation phase is entered by the formation of the stable TEC. Thus, position $n = 0$ does not correspond to the actual transcriptional starting point, but usually a few (8–10) nucleotides downstream. Finally, transcription will terminate at position $n = N$. Note that n is only affected by polymerization (lengthening) and depolymerization (shortening) of the nascent mRNA. The second variable m denotes the position of the polymerase's active site relative to n and ranges from $-n$ to 1. States $m = 0$ and $m = 1$ are defined as the pre- and posttranslocated states

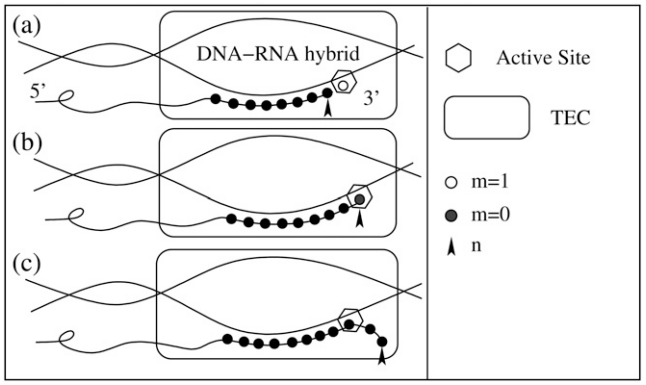
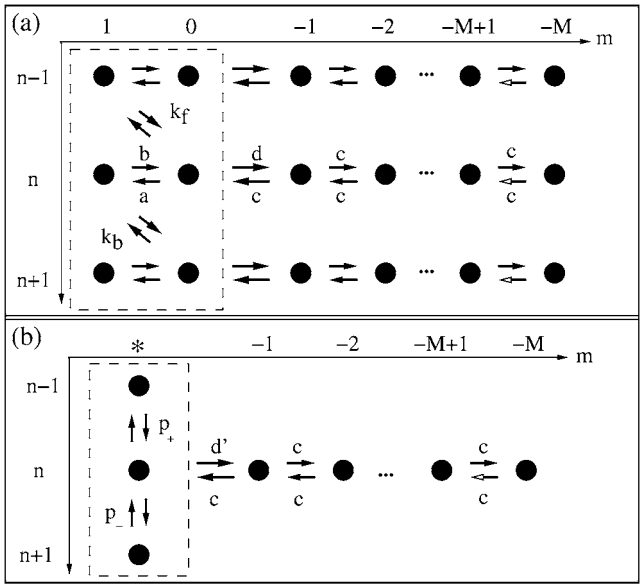


FIGURE 1 Schematic representation of the transcription elongation complex (TEC) at different translocation states: (*a*) Posttranslocated state at (n, $m = 1$), (*b*) pretranslocated state (n, $m = 0$), and (*c*) backtracked state (n, $m = -2$). The position of the TEC on the DNA template is characterized by the position of the active site, which in terms of variables n and m is $x = n + m$.



Parameter	Value	References
b/a	0.8	(61,62)
k_b/k_f	0.01	(26)
k_f	$36s^{-1}$	(63)

FIGURE 2 Schematic illustration of discrete models of transcription elongation. (*a*) Model A (*dotted rectangular*) includes polymerization, depolymerization, and transitions between the post- and pretranslocated states. Model B also allows for backward translocation of the TEC as far as $m = -M$, with $M \ll N$. If $n < M$, backward translocation is permitted up to state $m = -n$ (not shown). In the case of uninterrupted backtracking (case 1), the TEC can return from state $m = -M$ (*white arrow*), whereas in the case of transcript arrest (case 2), the TEC is halted at $m = -M$ until it is rescued by accessory factors, which move it to state ($n - M$, 0). The table includes typical values for parameters of Model A. (*b*) Schematic illustration of a simplified version of Models A and B when transition between pre- and posttranslocated states is the fastest process. The active states ($m = 0$, 1) have been collapsed into one state, denoted by the asterisk (*). At each template position the TEC can either proceed with polymerization, depolymerization, or enter a backtracked state, with effective rates p_+, p_-, or d', respectively.

of the TEC, respectively, while $m < 0$ corresponds to a backtracked (or reverse translocated) state. Hypertranslocation (which would lead to $m > 1$) is ignored.

The elongation phase starts with the TEC in state ($n = 0$, $m = 0$). The only transition possible from this state is to the posttranslocated state ($n = 0$, $m = 1$), from which the TEC can revert to ($n = 0$, $m = 0$) or proceed with polymerization. Polymerization, or the addition of a single nucleotide to the nascent mRNA strand, can only proceed from the posttranslocated state. Thus, with the TEC occupying the pretranslocated state (n, $m = 0$), polymerization by a single nucleotide requires two steps: 1), the TEC sliding forward to the posttranslocated state (n, $m = 1$); and 2), the extension of mRNA by one nucleotide ($n + 1$, $m = 0$), which leaves the TEC in the next pretranslocated state. Conversely, the reverse process of depolymerization can only proceed from the pretranslocated state and leaves the TEC in the previous posttranslocated state ($n - 1$, $m = 1$). Thus, at any given template position n, the TEC can freely move back and forth between the pretranslocated (n, $m = 0$) and the posttranslocated (n, $m = 1$) states, allowing depolymerization and polymerization, respectively, (except from the two boundary points $n = 0$ and $n = N$). A schematic diagram of state transitions for a simplified model excluding backtracking (Model A) is given in Fig. 2 *a*.

Inclusion of backtracking in the model provides an additional pathway, as the TEC can now hop from the pretranslocated state (n, $m = 0$) into the first backtracked state (n, $m = -1$). Subsequent backward translocation events can randomly shift the TEC's active site back and forth, possibly backtracking as far back as (n, $m = -n$) (8). In practice, backtracking is often restricted to $m = -M > -n$. In some cases, backtracking will consist of random reverse and forward translocations that eventually end as the TEC returns to the nucleotide target position (allowing polymerization to resume). In other instances, backtracking is interrupted (in a so-called transcript arrest (8)) and the pause eventually ends when the TEC is rescued by accessory factors, such as the Gre/TFIIS cleavage proteins (35,36). Note that backtracking affects only variable m, since it disrupts the positioning of the active site, leaving the length of the nascent mRNA (variable n) unaffected. In other words, both polymerization and depolymerization are blocked during backtracking until the corresponding target positions are recovered, i.e., (n, 1) and (n, 0), respectively. A schematic diagram of state transitions for a model of elongation with restricted backtracking (Model B) is given in Fig. 2 *a*.

For both Models A and B, we seek the statistics of the elongation time, i.e., the time needed for the TEC to reach position ($n = N$, $m = 0$) with the elongation phase starting with the TEC in state ($n = 0$, $m = 0$).

Model A: translocation-limited polymerization

In this model, backtracked states are ignored, and at each template position n only two translocation states are possible: $m = 1$ and $m = 0$, which allow transcript polymerization and depolymerization, respectively. The rates of polymerization and depolymerization are given by k_f and k_b, while a is the translocation rate from $m = 0$ to $m = 1$ and b the reverse rate from $m = 1$ to $m = 0$. (See typical values in the table of Fig. 2.)

The dynamics of $P_{n,m}(t)$, the probability of finding the polymerase in state (n, m) at time t, are described by the Master equation (37,38),

$$\frac{\partial P_{n,0}}{\partial t} = k_f P_{n-1,1} + bP_{n,1} - (k_b + a)P_{n,0}, \quad (1a)$$

$$\frac{\partial P_{n,1}}{\partial t} = k_b P_{n+1,0} + aP_{n,0} - (k_f + b)P_{n,1}, \quad (1b)$$

where n varies from 0 to $N - 1$. We assume that depolymerization is impossible at position $n = 0$ and that the process is terminated when position $n = N$ is reached. Consequently, the boundary conditions (BC) imposed on Eq. 1 should be reflecting at $n = 0$ and absorbing at $n = N$. The reflecting BC is obtained by defining a fictitious state $n = -1$ and setting $k_b P_{0,0} = k_f P_{-1,1}$. To obtain the absorbing BC, it is convenient to introduce a fictitious position at N and set $P_{N,0} = 0$ (38), which is equivalent to setting the transition rate from $(N - 1, 1)$ to $(N, 0)$ equal to zero.

A mean-field (quasi-steady-state) approximation yielding a biased random walk is obtained in the limit that the rates of polymerization are much slower than the rates of translocation (i.e., $k_f, k_b \ll a,b$) (26,28). The effective polymerization and depolymerization rates are $p_+ \approx k_f a/(a + b)$ and $p_- \approx k_b b/(a + b)$. We calculate μ, the mean elongation time (i.e., the time it takes for the TEC to arrive at $n = N, m = 0$ from a starting position at $n = 0, m = 0$) and the variance σ^2 as a function of the template length N (see Appendix A for a complete derivation). Under normal conditions, elongation is overwhelmingly favored over chain shortening (8) $K = p_-/p_+ \ll 1$. Therefore, we have

$$\mu = \frac{N}{p_+} + K\frac{(N-1)}{p_+} + \mathcal{O}(K^2), \quad (2a)$$

$$\sigma^2 = \frac{N}{p_+^2} + K\frac{(4N-4)}{p_+^2} + \mathcal{O}(K^2). \quad (2b)$$

Fig. 3 shows results obtained from stochastic simulations of Model A (Eq. 1), along with the analytic results obtained

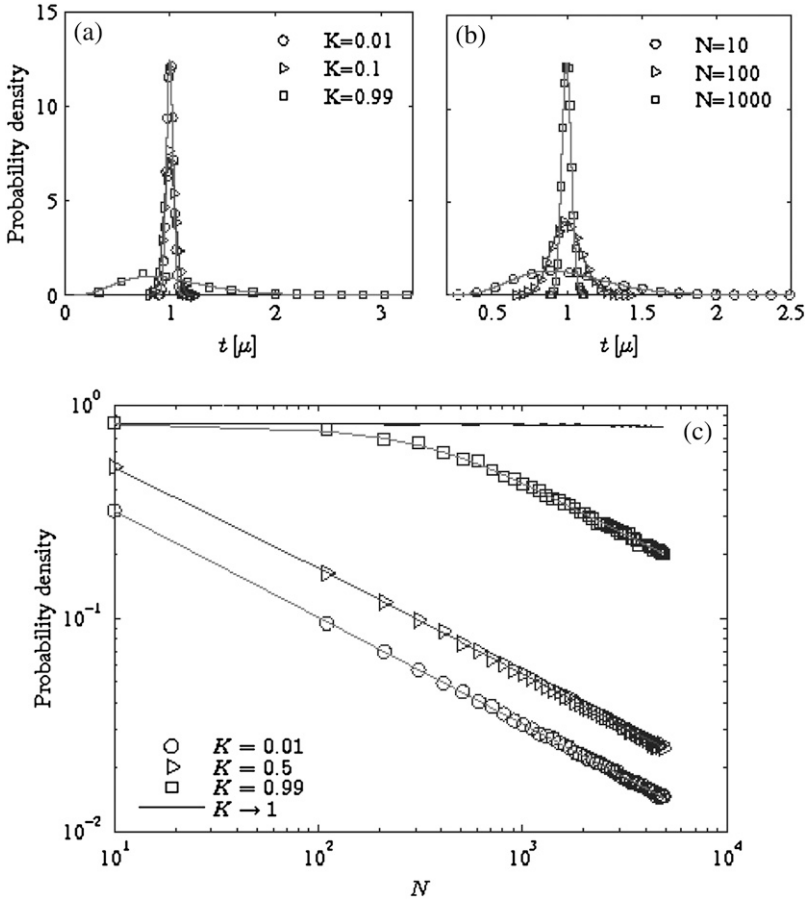


FIGURE 3 (*a*, *b*) Distribution of dimensionless elongation times (scaled by the mean elongation time) for Model A (Eq. 1). Mean-field analytic results are plotted in solid curves, and superimposed with stochastic simulations results. (*a*) Results for $N = 1000$ bp, $p_+ = 20\ s^{-1}$ and different polymerization biases $K = 0.01, 0.5, 0.99$. (*b*) Results for $K = 0.01$, $p_+ = 20\ s^{-1}$ and different template lengths $N = 10, 100, 1000$ bp. (*c*) Standard deviation over mean (σ/μ) plotted against the template length N for different values of K. As expected, the width of the distribution scales as $1/\sqrt{N}$.

in the mean-field approximation, for different values of N and K. In the small K regime and for small values of N, the elongation times are approximately γ-distributed, with shape parameter $\alpha = \mu^2/\sigma^2$ and scale parameter $\beta = \sigma^2/\mu^2$. As N is increased, the distribution approaches a Gaussian, in agreement with the Central Limit Theorem, with mean and variance given by Eqs. 2a and 2b, respectively. Since both μ and σ^2 scale linearly with the template length N, fluctuations around the mean are of the order $1/\sqrt{N}$. As a result, the distribution becomes narrowly peaked around the mean as N is increased, and in the limit $N \to \infty$, where fluctuations tend to zero, the process becomes essentially deterministic. Conversely, in the $K \to 1$ limit, polymerization and depolymerization tend to play equal roles, leading to fluctuations in the transcription time that do not vanish as N is increased.

Model B: transcription with backtracking pauses

We now extend Model A to include elongation pauses that arise when the TEC occupies backtracked states ($m < 0$). In particular, a pause is signaled when the TEC enters the backtracked state $m = -1$ from state $m = 0$. We denote the corresponding transition rate by d and assume a slow rate relative to polymerization $d \ll p_+$. From $m = -1$ the TEC hops across contiguous backtracked states with rate c. In principle, at each template position n, backtracking can proceed up to $m = -n$ (8). However, in practice, different mechanisms, such as RNA hairpins, RNA-DNA interactions, and cleavage enzymes preclude extensive backtracking (33). A more reasonable assumption is that backtracking is restricted in length; we assume backtracking to be restricted to a fixed number of steps $m = -M \gg -n$, which we take to be independent of position n. Also, for values of template position n that are smaller than M, backtracking is permitted to extend as far as $m = -n$. In fact, hairpins are dynamic (breaking and reforming), implying that the choice of fixed M is only a first approximation. If the hairpin relaxation time is sufficiently fast (as compared with the backtracking rate), such dynamics could lead to fluctuations in the value of M.

Dynamics of backtracking pauses

To gain insight into the statistics of transcriptional pauses, we describe and examine the dynamics of backtracking as a separate process. Without loss of generality, we describe backtracking by a symmetric hopping process, or unbiased random walk with rate c. The asymmetric case, equivalent to a biased random walk, is quite a straightforward generalization (39). For simplicity, we characterize backtracked states by a new variable $l = -m$, where $1 \le l \le M$. The probability $P(l, t)$, of finding the polymerase in state l at time t given that it was in state $l = 1$ at $t = 0$, follows the Master equation:

$$\frac{\partial P(l,t)}{\partial t} = cP(l-1,t) + cP(l+1,t) - 2cP(l,t). \quad (3)$$

We use the Laplace transform, $\tilde{p}(l,s) = \int_0^\infty P(l,t)e^{-st}dt$, to obtain exact expressions for the probability distribution of the duration of backtracking pauses for two different scenarios:

1. Uninterrupted backtracking: $l = M$ is a reflecting boundary, and termination of the pause occurs when the TEC eventually slides back to state $l = 0$ and
2. Transcript arrest: The TEC is irreversibly halted at $l = M$. Elongation can be resumed either from state $l = 0$ or from position $l = M$ with the aid of accessory factors. Detailed derivations are given in Appendix B.

Case 1: uninterrupted backtracking

In this case no backward translocation is possible beyond state $l = M$, and the pause is ended when state $l = 0$ is reached. The corresponding boundary conditions for Eq. 3 are: $P(0, t) = 0$ (absorbing) and $cP(M, t) = cP(M + 1, t)$ (reflecting). The mean pause duration is $\langle t \rangle = M/c$ and an analytic expression for the probability distribution $\mathcal{P}(t)$ of pause duration is given in Appendix B. Simple expressions for $\mathcal{P}(t)$ are obtained in the following limits:

$$\mathcal{P}(t) \approx \begin{cases} \dfrac{1}{2\sqrt{\pi}\sqrt{c}t^{3/2}}, & \dfrac{1}{c} \ll t \ll \dfrac{M^2}{c}, \\ \dfrac{\pi c \sin\left(\dfrac{\pi}{2(M+1)}\right)}{(1+M)^2} \exp\left[e^{-\dfrac{c\pi^2}{4(1+M)^2}t}\right], & t \gg \dfrac{M^2}{c}. \end{cases} \quad (4)$$

For times short compared to the timescale of diffusion to the reflecting state $l = M$ ($t \ll M^2/c$), but still longer than the time for the TEC to diffuse by one nucleotide ($t \gg 1/c$), $\mathcal{P}(t)$ scales as $t^{-3/2}$. Interestingly, the power law behavior characteristic of this regime is consistent with the heavily skewed and heavy-tailed distribution observed by Shaevitz et al. (19). Conversely, for times much longer than M^2/c, which ensure reflection, the asymptotics are altered and $\mathcal{P}(t)$ exhibits a rapid exponential decay. The two different asymptotic behaviors are illustrated in Fig. 4 *a*, where the analytic results have been plotted together with the data obtained from stochastic simulations of the model.

Case 2: backtracking with transcript arrest

As before, pauses begin with a transition into state $l = 1$ and terminate when state $l = 0$ is reached. However, in this scenario, backtracking will also be terminated by the arrest of transcription if the TEC arrives at $l = M$. Transcription can only resume from the arrested state with the aid of a rescue mechanism (35,36). The boundary conditions imposed to Eq. 3 are therefore absorbing at both ends: $P(0, t) = P(M, t) = 0$.

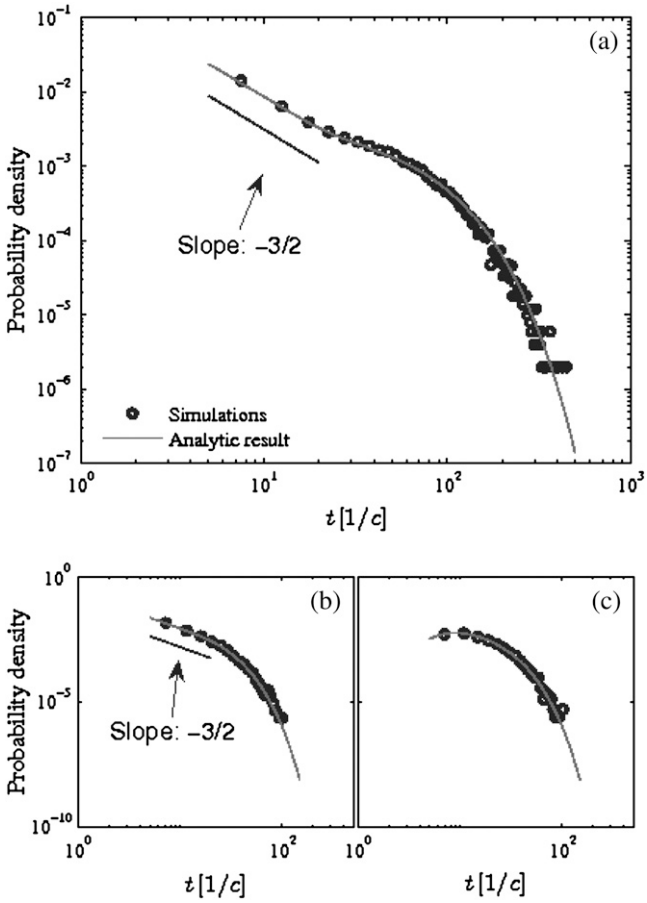


FIGURE 4 Results for case 1 (uninterrupted backtracking) and case 2 (transcript arrest) pauses with $M = 10$. Distributions of (a) pause duration $\mathcal{P}(t)$ for case-1; (b) self-recovered pause duration $\mathcal{P}_0(t)$ for case-2; and (c) time to arrest $\mathcal{P}_M(t)$ for case-2. Plotted are the analytic results (Eq. 39, and Eqs. 45a and 45b, respectively) as solid lines and the results of stochastic simulations as circles. $\mathcal{P}(t)$ and $\mathcal{P}_0(t)$ exhibit a power law decay for $1/c \ll t \ll M^2/c$, followed by an exponential cutoff in long time limit ($t \gg M^2/c$).

It can be shown (see Appendix B) that the probability of eventual arrest of the TEC is $p_M = 1/M$; the probability of TEC recovery from the pause is $p_0 = 1-p_M$; and the corresponding mean time for each case is $\langle t\rangle_M = (M^2 - 1)/6c$ and $\langle t\rangle_0 = (2M - 1)/6c$. Compact expressions for $\mathcal{P}_0(t)$, the probability distribution of recovering from the pause at time t, are obtained in the two limits discussed above:

$$\mathcal{P}_0(t) \approx \begin{cases} \dfrac{1}{2\sqrt{\pi}\sqrt{c}t^{3/2}}, & \dfrac{1}{c} \ll t \ll \dfrac{M^2}{c}, \\ \dfrac{2\pi c \sin\left(\dfrac{\pi}{M}\right)}{M^2} \exp\left[e^{-\dfrac{\pi^2 c}{M^2}t}\right], & t \gg \dfrac{M^2}{c}. \end{cases} \quad (5)$$

Once again, the distribution demonstrates a power law decay for $1/c \ll t \ll M^2/c$, followed by an exponential cutoff. For sufficiently long times $t \gg M^2/c$ that allow diffusion to the boundary $l = M$, the probability distribution of the TEC arrest decays exponentially with $\mathcal{P}_M(t) \approx \mathcal{P}_0(t)$. The above analytic results, along with stochastic simulations, are summarized in Fig. 4.

Stochastic simulations of Model B

Having characterized backtracking statistics, we are now in a position to examine the effects of backtracking on the total elongation time. The macroscopic (observable) properties that we must consider are: 1), the number of pauses δ over a DNA template of length N, and 2), the aggregate lifetime of all the pauses relative to the time spent on active polymerization. These properties are linked to the microscopic parameters d and c, respectively. In particular, when translocation between pre- and posttranslocated states is the fastest process, the number of pauses δ is given by:

$$\frac{\delta}{N} = \frac{d\dfrac{a}{a+b}}{d\dfrac{a}{a+b} + p_+ + p_-} = \frac{d'}{d' + p_+ + p_-}, \quad (6)$$

where $d' = d(a/(a+b))$ is the effective rate of entering into a backtracked state (see Fig. 2 b). Moreover, the distribution of pause durations (for the case of uninterrupted backtracking) is determined by the symmetric diffusion rate c, with M/c being the mean pause duration.

As expected, in the limit of short-lived pauses, even the aggregate pause duration will be negligible relative to the time spent on processive polymerization, $N/p_+ \gg \delta(M/c)$, and so the distribution of elongation times will approach that of Model A. Conversely, when $N/p_+ \gg \delta(M/c)$, pauses dominate the total elongation time and the distribution of elongation times is significantly affected by the large fluctuations in the duration of the pauses. In the limit $p_+ \gg d'$ and $p_+ \gg p_-$, Eq. 6 becomes $\delta/N \approx d'/p_+$ and the above limits can be written as $d'(M/c) \ll 1$ and $d'(M/c) \gg 1$. We therefore introduce $R = d'(M/c)$ as a dimensionless measure of pauses which quantifies their relative contribution to the elongation time. This measure of pause durations is particularly useful as it is directly linked to the macroscopic parameters of the system (i.e., mRNA production rate) but is derived from the microscopic rate parameters.

Figs. 5 and 6 illustrate the results of the stochastic simulations of Model B, i.e., transcription with restricted, uninterrupted backtracking, for different values of R (keeping the frequency of pauses δ/N constant). As expected, for $R \to 0$ the polymerization-only model (Model A) is recovered and $\sigma/\mu = 1/\sqrt{N}$ (Fig. 6). This is also evident from the distribution of elongation times, where for small R the high peak close to the mean elongation time predicted by Model A indicates that either no pauses or only brief ones occur. The effect of backtracking events is most evident in the heavier tail of the distribution since rare prolonged pauses can give rise to significantly longer elongation times. This effect is magnified as the fraction of time spent in pauses is increased (i.e., for higher values of R) (Fig. 5 a). For increasing pause frequency

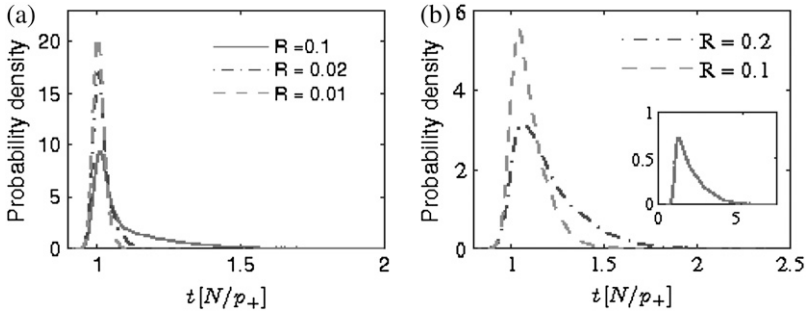


FIGURE 5 Distributions of dimensionless elongation times (scaled by N/p_+) for Model B for different values of $R = d'M/c$. The distributions were obtained from stochastic simulations. (*a*) $N = 4$ kb, $M = 10$ bp, $p_+ = 10\ s^{-1}$, $K = 0.01$ and d' chosen to yield $\delta/N \approx d'/p_+ = 1$ pauses/kb (19,22). (*b*) $N = 1$ kb, $M = 10$ bp, $p_+ = 10\ s^{-1}$, $K = 0.01$, and d' chosen to yield $\delta/N \approx d'/p_+ = 10$ pauses/kb. (*Inset*) $R = 1$. The effect of the pauses is evident in the heavy tails that broaden with decreasing R or increasing δ/N.

(higher δ/N) the effect on the total elongation time is clearly more profound; the distribution becomes broader and exhibits a general shift toward longer elongation times (Fig. 5 *b*).

mRNA transcript levels: production and degradation

Models A and B capture the statistics of the elongation phase. Ultimately, however, one is interested in the mRNA levels, which are the combined action of mRNA production (transcription) and degradation. In general, the transcription process involves an initiation phase (which includes promoter binding, open complex formation, and promoter clearance), an elongation phase, and termination. As a more complete model of transcription, we assume fast termination and combine the model of elongation presented above (Model B) with a simplified, first-order initiation step. Degradation is also represented as a Poisson, single-step process. Using

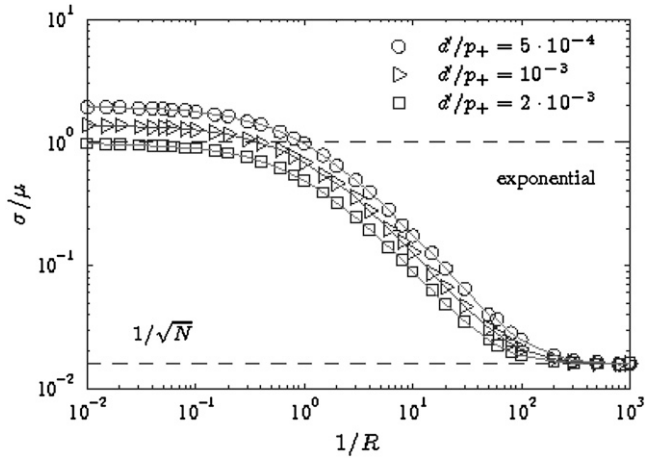


FIGURE 6 Standard deviation over mean (σ/μ) of elongation times (Model B) plotted against $1/R$ for different values of the ratio d'/p_+ (pause frequency). As $1/R \to 0$, pauses become more significant and the distribution of elongation times becomes broader. In the case of frequent pausing ($d'/p_+ = 2 \times 10^{-3}$), the distribution exhibits characteristics of an exponential distribution, i.e., $\sigma/\mu = 1$ (indicated by the *upper dashed line*). As $1/R \to \infty$, the effect of pauses vanishes and Model B approaches Model A, where $\sigma/\mu \approx 1/\sqrt{N}$ (indicated by the *lower dashed line*). Parameters used: $N = 4$ kb, $M = 10$ bp, $d' = 0.01\ s^{-1}$, $K = 0.01$, and $p_+ = 2$, 10, and 20 s^{-1}.

stochastic simulations of this combined transcription-degradation model, we examine how the elongation and possible pauses therein affect steady-state mRNA levels.

We denote the initiation rate as k_i. The elongation phase proceeds as described by Model B and instantaneous termination takes place when the transcript reaches its designated size, leading to mRNA production. Finally, mRNA degradation is modeled as a first-order process with rate constant k_d. The combination of mRNA production and degradation gives a first handle on mRNA levels and fluctuations in the cell.

In fact, mRNA production is complicated by the fact that multiple initiation events can occur within the time it takes to produce a single mRNA. This would lead to several TECs moving in tandem on the same DNA template (40), each synthesizing a nascent mRNA. To capture the fact that two TECs cannot come in close proximity due to nonspecific interactions between them or to the additional work required to deform the DNA helix (41,42), we set a minimum (exclusion) distance of L nucleotides ($L \ll N$) between the active sites of any two contiguous TECs. In terms of variables n and m of Model B, the active site of a TEC is located at position $x = n + m$ along the DNA template. Therefore, a TEC, positioned at x_1, can translocate forward (backward) if the leading (trailing) TEC, positioned at x_2, is at distance of at least L nucleotides, i.e., $|x_1 - x_2| < L$. A similar argument applies for transcription initiation, that is, no RNA polymerase can initiate transcription if a TEC is at position $x \leq L$. A schematic illustration of the model is given in Fig. 7.

The relevant timescales associated with the above model are: 1), the time needed for transcription initiation $\tau_1 = 1/k_i$; 2), the time needed by the TEC to transcribe L nucleotides $\tau_2 \approx L/p_+$; and 3), the mean time of a pause due to backtracking $\tau_3 = M/c$. When initiation is the rate-limiting step ($\tau_1 \gg \tau_2,\tau_3$), the density of TECs on the DNA template is low and therefore transcriptional pauses and interactions between TECs are expected to have marginal effects. Consequently, the rate of mRNA production is set mainly by the rate of initiation k_i and the statistics of the mRNA levels are expected to be approximately Poisson with the mean equal to the variance ($\mu_{mRNA} = \sigma^2_{mRNA}$; see Fig. 8 *III*). If the rate of polymerization is the rate-limiting step

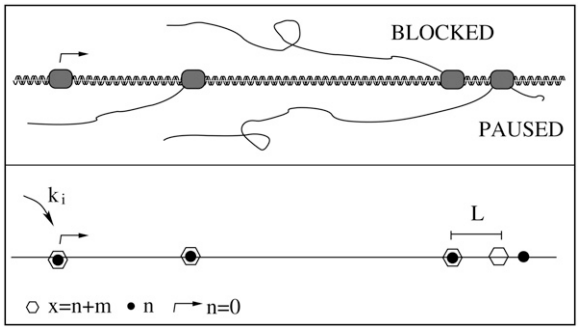


FIGURE 7 Schematic illustration of multiple RNAP molecules simultaneously transcribing a DNA template. Transcription initiation proceeds with an effective rate of k_i. The position of each TEC on the DNA is characterized by the position of its active site, which is given by $x = n + m$. We also set a minimum (exclusion) distance of L nucleotides between any two TECs. If transcriptional pauses are sufficiently long they can block the progress of trailing RNAP molecules and subsequently lead to a burst in mRNA production. Such a scenario suggests a significant link between transcriptional pauses and mRNA production statistics.

($\tau_2 \gg \tau_1,\tau_3$), fast transcription initiation is blocked by the slow movement of the TECs on the DNA template, while the relatively short-lived backtracking events, as in the case above, play no significant role. In particular, the density of TECs along the DNA is expected to be maximal (N/L), with the TECs kept evenly spaced (L nucleotides apart) by exclusive interactions. In this regime the statistics of the mRNA levels are sub-Poisson with more evenly distributed TECs along the DNA template ($\mu_{\mathrm{mRNA}} > \sigma^2_{\mathrm{mRNA}}$; see Fig. 8 *II*). Finally, $\tau_3 \gg \tau_1,\tau_2$ corresponds to a regime where long pauses dominate transcription. Such pauses can create congestion points by blocking the movement of trailing TECs, while the leading TECs continue to transcribe normally. In this way the uniform ($\tau_2 \gg \tau_1$) or Poisson ($\tau_1 \gg \tau_2$) distribution of TECs on the DNA template is disrupted, resulting in a burstlike production of mRNA transcripts (Fig. 9) and super-Poisson mRNA statistics (i.e., $\mu_{\mathrm{mRNA}} < \sigma^2_{\mathrm{mRNA}}$; see Fig. 8 *I*).

In the bursting regime, the effect of elongation pauses can be linked heuristically to a switching mechanism between high and low rates of mRNA production. In particular, sufficiently long pauses shut down mRNA production by blocking trailing TECs. In the intervals between pauses, multiple blocked TECs that have accumulated at a congestion site are likely to be transcribed in a burst of rapid mRNA production. A qualitative description of the different classes of behavior obtained for the integrated initiation, elongation, degradation model is presented in Table 1. Stochastic simulations of the model confirm that rare and long-lived pauses give rise to jamming of TEC trafficking during transcription and therefore bursts of mRNA production. We note that such abrupt switching between two states is reminiscent of dynamic phenomena observed in studies of the asymmetric exclusion process (43,44).

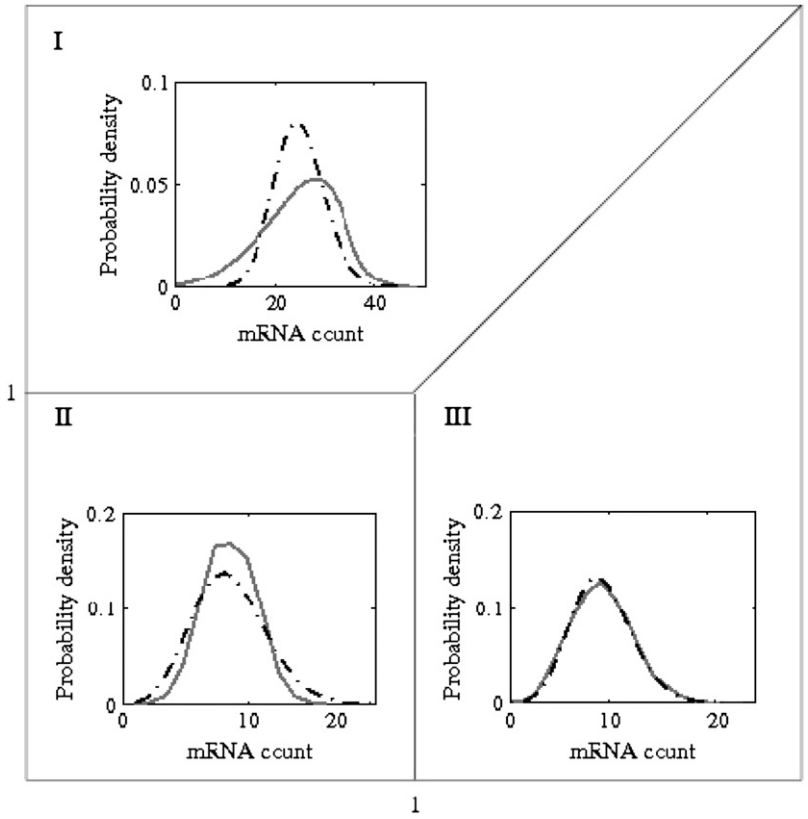


FIGURE 8 Distribution of steady-state number of mRNA molecules (*solid line*). Simulations included transcription initiation, elongation, and mRNA degradation and allowed multiple RNAP molecules to transcribe the DNA template at the same time. A Poisson distribution with the same mean value is given for reference (*dash-dotted line*). (*I*) When elongation pauses are longer than the time needed for transcription initiation and the time needed by the TEC to transcribe L nucleotides ($\tau_3 \gg \tau_1,\tau_2$), the mRNA distribution is expected to be broader than Poisson. (*II*) When the movement of RNAP molecules on the DNA template is the rate-limiting step ($\tau_2 \gg \tau_1,\tau_2$), the mRNA distribution predicted by the model is sub-Poisson. (*III*) When transcription initiation is the rate-limiting step ($\tau_1 \gg \tau_2,\tau_3$), the mRNA distribution predicted by the model is Poisson.

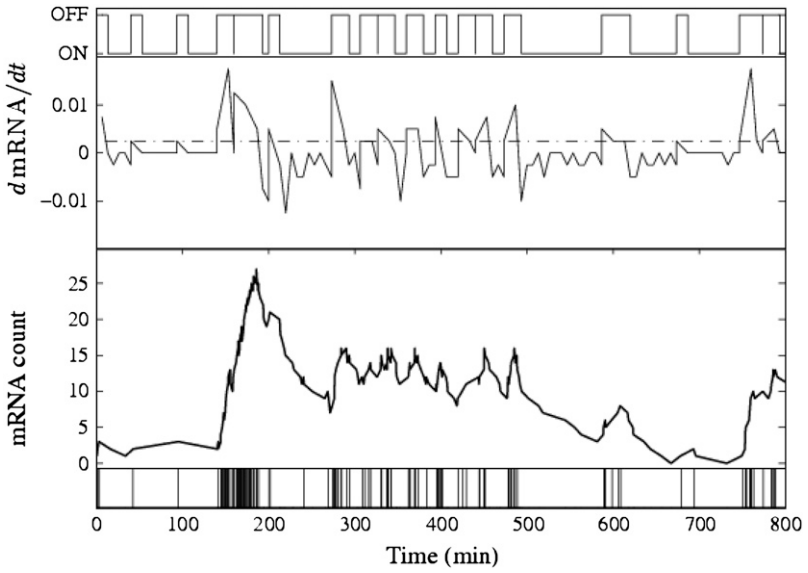


FIGURE 9 Simulation of mRNA population levels in an integrated model of transcription initiation, elongation, and mRNA degradation (parameters given in Appendix C; 10^3 runs). The inclusion of transcriptional pausing (when multiple initiations are permitted) results in bursts of mRNA production and super-Poisson mRNA statistics ($\sigma^2_{\mathrm{mRNA}}/\mu_{\mathrm{mRNA}} = 4.25$). The bottom panel shows the mRNA production events in time and the trace above illustrates the resulting mRNA count fluctuations. In the third panel, dmRNA/dt is plotted ($dt = 6$ min), along with an arbitrary threshold (*dotted line*, set to 1/dt mRNA/s). The threshold enables us to visualize the transcriptional process as a telegraph process with off- and on-states corresponding to low and high rates of mRNA production (*top panel*).

DISCUSSION

We have presented a stochastic model of transcription, including initiation, elongation, and mRNA degradation. Our main focus has been on the elongation phase for which we obtained analytic results both for the polymerization dynamics (ignoring backtracking) and for the dynamics of backtracking pauses. Our model of backtracking pauses as a first passage process is consistent with recent single molecule experiments (19). By means of stochastic simulations we have also examined how pauses affect the total elongation times. Finally, we have developed a model of mRNA production and degradation that combines transcription initiation, transcription elongation, and mRNA degradation. In this model, multiple RNAPs with repulsive interactions can move in tandem on the same DNA template. We used stochastic simulations of this model to examine how the dynamics of the elongation phase and backtracking pauses therein affect the statistics of the mRNA population levels.

Our key results are particularly instructive in two limits: first, when pauses cause a weak perturbation to elongation dynamics and secondly, when they significantly affect it. The third regime, in which initiation is the rate-limiting step (with relatively rapid elongation), recovers previously predicted Poisson statistics. As expected, if the elongation phase dominates transcription, but the time spent in backtracking pauses is brief relative to that spent on active polymerization, similar results to the polymerization-only model are recovered. That is, for sufficiently long sequences ($N \gg 1$) the elongation times follow a narrow Gaussian distribution with fluctuations around the mean scaling like $1/\sqrt{N}$, where N is the length of the gene. This leads to a characteristic delay in the total time of transcription. Coupling fast transcriptional initiation with such a model of transcription elongation predicts a more homogeneous transcription process and hence steadier mRNA population levels than would be produced by a model of initiation alone.

TABLE 1 Table summarizing the behavior of mRNA production in the different limiting regimes (with time-limiting initiation, polymerization, or pausing kinetics)

	Regime	Behavior
$\tau_1 \gg \tau_2,\tau_3$	$\tau_1 \gg \tau_2,\tau_3$	Poisson
	$\tau_1 \gg \tau_2,\tau_3$	Poisson
$\tau_2 \gg \tau_1,\tau_3$	$\tau_1 \gg \tau_3$	sub-Poisson
	$\tau_3 \gg \tau_1$	sub-Poisson
$\tau_3 \gg \tau_1,\tau_2$	$\tau_1 \gg \tau_2$	super-Poisson
	$\tau_1 \gg \tau_2$	super-Poisson
$\tau_1 - \tau_2 \gg \tau_3$		sub-Poisson
$\tau_1 - \tau_3 \gg \tau_2$		super-Poisson
$\tau_2 - \tau_3 \gg \tau_1$		super-Poisson

In the opposite regime, when there is a significant number of backtracking pauses whose duration is comparable to the active polymerization time, there is a dramatic change in the distribution of transcriptional times. We considered two types of backtracking pauses; pauses that end with the TEC sliding back into position and backtracking pauses that can lead to transcriptional arrest. For both classes of pauses we found a broad distribution of pause durations with a power law decay cutoff by an exponential one. Consequently, the statistics of the elongation phase can be dramatically altered, with increased mean and a significantly broader distribution of elongation times, which mirrors the distribution of pause durations.

Recent experiments have provided evidence for the existence of bursts of transcription both in bacterial (11) and eukaryotic cells (30,31). We have found that our model of the dynamics of elongation with pauses leads naturally to switching between high and low mRNA production rates, resulting in transcriptional bursts. Our findings suggest that rare and long elongation pauses (from the tails of the distribution) act as congestion points turning off mRNA

production for long periods, while allowing rapid mRNA production for short intervals. Such long pauses, therefore, give rise to more strongly fluctuating mRNA levels. Thus, in this regime, elongation pauses act as a rate-limiting step.

In fact, experimental reports of transcriptional bursting measure mRNA population levels (rather than production rates). We obtain consistent fluctuations in mRNA population levels, in a model that combines transcription with mRNA degradation kinetics. Other possible elongation pauses (which are not linked to backtracking) could result in similar bursting effects (45). Indeed, pauses can, in general, result from sequence-encoded signals (46), elongation factors, or nucleosome packaging (47,48). We note, however, that the rate-limiting step can also be provided by a number of different mechanisms associated with the transcription process, such as changes in the state of the promoter (30,31) (e.g., by chromatin remodeling) or the diffusive motion of regulatory molecules (49).

While single molecule studies have provided evidence that RNAP backtracking dominates in vitro transcription and results in pauses of significant (>20 s) duration (19), it is interesting to consider how frequent they are and what role they may play in vivo. For example, backtracking pauses have been previously implicated in mRNA editing and error correction (8,23) and could therefore partially account for discrepancies between theoretically expected and observed error rates in mRNA transcripts. Differences in free energies between correct and incorrect nucleotides yield an expected error rate of 10^{-3} errors/bp. This high rate contrasts with experimentally measured values of 10^{-5} errors/bp (50). This discrepancy in error rates could presumably be accounted for by error correction mechanisms, which may include backtracking pauses (M. Voliotis, N. Cohen, C. Molina-París, and T. B. Liverpool, unpublished). Of course, the situation in vivo is further complicated by the effects of transcription factors and other regulatory proteins. Nevertheless, if backtracking pauses are significant in the elongation process they could provide the cell with ample opportunity for a range of regulation mechanisms.

The models presented here relied on a number of simplifying assumptions. In particular, both polymerization and elongation pauses were taken to be sequence-independent. The assumption that polymerization takes place on a homogeneous DNA template is likely to be a simplification, since the local rates of translocation have been suggested to depend on the underlying local DNA sequence. Moreover, our models have neglected any sequence dependence that has been attributed to short-lived pauses (20,21). We leave the development of more detailed sequence-dependent kinetic models of elongation dynamics for future research.

While in this article we restrict our calculations to models of transcription, similar arguments regarding pauses and bursting should also be relevant for translation. Applications of these results will ultimately contribute to a more complete understanding of gene expression and regulation, and fluctuations therein. A better understanding of these processes will also shed light on the differences between the effects of gene regulatory mechanisms, which act during transcription and translation (18,52–56) as compared to those which act by controlling the initiation of these processes. Ultimately, models of noise generation in the cellular environment may lead to new insights on the ways in which cells survive and adapt, with consequences for cell development, function, and fate.

APPENDIX A: TRANSLOCATION-LIMITED POLYMERIZATION

For Model A, the Master equation describing the dynamics of $P_{n,m}(t)$, the probability of finding the TEC in state (n, m) at time t, starting from an initial state (0, 0) at $t = 0$, is given by Eq. 1. Since we take N to be the termination site, we implement an absorbing boundary at position $(n = N, m = 0)$. Such a boundary can in general be obtained by setting the depolymerization rate at $n = N$ equal to 0. By doing so, Eq. 1b is affected only for $(n = N - 1, m = 1)$:

$$\frac{\partial P_{N-1,1}}{\partial t} = aP_{N-1,0} - (k_f + b)P_{N-1,1}. \quad (7)$$

The same result can be obtained by setting $P_{N,0} = 0$ and regarding Eq. 1b valid for every n in $\{0, 1, \ldots, N-1\}$. Also, since we assume $(n = 0, m = 0)$ to be a reflecting boundary, we set the depolymerization rate at $n = 0$ to 0 and $P_{-1,1} = 0$, i.e., there is no probability flow from or to state $(n = -1, m = 1)$. In this way, Eq. 1a is affected only for $(n = 0, m = 0)$:

$$\frac{\partial P_{0,0}}{\partial t} = bP_{0,1} - aP_{0,0}. \quad (8)$$

The same result can be obtained by setting $k_b P_{0,0} = k_f P_{-1,1}$ such that Eq. 1a is valid for every n in $\{0, 1, \ldots, N-1\}$.

We can define a mean occupancy for each translocation state $(m = 0, 1)$ by summing over all possible template positions, $\Pi_m(t) = \sum_{n=0}^{N-1} P_{n,m}(t)$. From Eq. 1a, we obtain

$$\frac{\partial \Pi_0}{\partial t} = (k_f + b)\Pi_1 - (k_b + a)\Pi_0, \quad \text{and} \quad \Pi_1 = 1 - \Pi_0. \quad (9)$$

The solution to Eq. 9 that satisfies initial conditions $\Pi_0(0) = 1$ relaxes on a timescale $\tau = (a+b+k_f+k_b)^{-1} \ll k_f^{-1}$. On timescales longer than τ, this solution attains steady-state values such that $\Pi_0^s = (k_f + b)\tau$ and $\Pi_1^s = (k_b + a)\tau$. For such long times the fluctuations in n and m become independent and we can write $P_{m,n} = \Pi_m^s P_n$. Substituting back into Eq. 1 and summing over m, we obtain

$$\frac{\partial \mathcal{P}_n}{\partial t} = p_- \mathcal{P}_{n+1} + p_+ \mathcal{P}_{n-1} - (p_- + p_+)\mathcal{P}_n, \quad (10)$$

which is equivalent to a biased random walk with effective polymerization and depolymerization rates

$$p_+ = k_f(k_b + a)\tau \approx \frac{k_f a}{a + b}, \quad (11a)$$

$$p_- = k_b(k_f + b)\tau \approx \frac{k_b b}{a + b}, \quad (11b)$$

where we have used $k_f, k_b \ll a,b$. Note that the boundary conditions for Eq. 10 are $\mathcal{P}_N = 0$ (absorbing) and $p_-\mathcal{P}_0 = p_+\mathcal{P}_{-1}$ (reflecting).

The elongation time is defined as the time needed for the TEC to reach position $(n = N, m = 0)$ starting from $(n = 0, m = 0)$. In the mean-field model the mean and variance of the elongation time can be calculated using the backward Master equation (38). We denote the initial template position

of the TEC at time $t_0 = 0$ by n_0 and rewrite Eq. 10 in terms of conditional probabilities:

$$\frac{\partial \mathcal{P}(n,t|n_0,t_0)}{\partial t} = p_+ \mathcal{P}(n-1,t|n_0,t_0) + p_- \mathcal{P}(n+1,t|n_0,t_0) - (p_+ + p_-)\mathcal{P}(n,t|n_0,t_0). \tag{12}$$

The backward Master equation is (38)

$$\frac{\partial \mathcal{P}(n,t|n_0,t_0)}{\partial t_0} = p_+ [\mathcal{P}(n,t|n_0,t_0) - \mathcal{P}(n,t|n_0+1,t_0)] + p_- [\mathcal{P}(n,t|n_0,t_0) - \mathcal{P}(n,t|n_0-1,t_0)]. \tag{13}$$

Since the system is homogeneous, we can write

$$\mathcal{P}(n,t|n_0,t_0=0) = \mathcal{P}(n,0|n_0,-t), \tag{14}$$

so that the backward Master equation takes the form

$$\frac{\partial \mathcal{P}(n,t|n_0,0)}{\partial t} = p_+ [\mathcal{P}(n,t|n_0+1,0) - \mathcal{P}(n,t|n_0,0)] + p_- [\mathcal{P}(n,t|n_0-1,0) - \mathcal{P}(n,t|n_0,0)]. \tag{15}$$

The boundary conditions for the backward Master equation are $\mathcal{P}(n,t|n_0 = 0,0) = \mathcal{P}(n,t|n_0 = -1,0)$ (reflecting) and $\mathrm{P}(n,t|n_0 = N,0) = 0$ (absorbing).

The probability that at time t the TEC has not yet reached the absorbing boundary is given by

$$\sum_{n=0}^{N-1} \mathcal{P}(n,t|n_0,0) = G(n_0,t). \tag{16}$$

If T is the elongation time (time needed to complete elongation by reaching the absorbing boundary at position $n = N$), $G(n_0, t)$ is the probability that $T \geq t$. In other words, the cumulative distribution function of the elongation times is $1 - G(n_0, t)$. We sum Eq. 15 over n from $n = 0$ to $n = N - 1$ to obtain

$$\frac{\partial G(n_0,t)}{\partial t} = p_+ [G(n_0+1,t) - G(n_0,t)] + p_- [G(n_0-1,t) + G(n_0,t)], \tag{17}$$

subject to the initial condition $G(n_0,0) = 1$ and boundary conditions $G(N,t) = 0$ and $G(0,t) = G(-1,t)$.

Equation 17 can be expressed and solved in terms of the first and second moments of the elongation time T, which can be written as

$$T(n_0) = \langle T \rangle = -\int_0^{+\infty} t \partial_t G(n_0,t) dt = \int_0^{+\infty} G(n_0,t) dt, \tag{18}$$

$$T_2(n_0) = \langle T^2 \rangle = -\int_0^{+\infty} t^2 \partial_t G(n_0,t) dt = 2\int_0^{+\infty} t G(n_0,t) dt. \tag{19}$$

We integrate Eq. 17 with respect to t to obtain

$$-1 = p_+ T(n_0+1) + p_- T(n_0-1) - (p_+ + p_-)T(n_0) = p_+ [T(n_0+1) - T(n_0)] + p_- [T(n_0-1) - T(n_0)]. \tag{20}$$

The boundary conditions imply $T(N) = 0$, $T(0) = T(-1)$. To solve this difference equation we introduce

$$U(n_0) = T(n_0) - T(n_0-1), \tag{21}$$

and substituting into Eq. 20 yields

$$p_+ U(n_0+1) - p_- U(n_0) = -1. \tag{22}$$

Solving the above two difference equations recursively, we obtain (38)

$$T(n_0) = \sum_{n=n_0+1}^{N} \frac{1}{p_+} \sum_{n'=0}^{n-1} \left(\frac{p_-}{p_+}\right)^{n'}. \tag{23}$$

By setting $K = \mathrm{p}_-/\mathrm{p}_+$ and observing that $0 \leq K < 1$, we can write

$$T(n_0) = \frac{1}{p_+} \sum_{n=n_0+1}^{N} \frac{1-K^n}{1-K} = \frac{1}{p_+(1-K)} \left[N - n_0 - \frac{K^{n_0+1} - K^{N+1}}{1-K} \right]. \tag{24}$$

Finally by letting $n_0 = 0$, we obtain the mean elongation time

$$\mu = \frac{1}{p_+(1-K)} \left[N - \frac{K(1-K^N)}{1-K} \right]. \tag{25}$$

For the variance of the elongation time we carry out a similar derivational. Multiplying by t and integrating Eq. 17 over t, we obtain

$$-2T(n_0) = p_+ T_2(n_0+1) + p_- T_2(n_0-1) - (p_+ + p_-)T_2(n_0) = p_+ [T_2(n_0+1) - T_2(n_0)] + p_- [T_2(n_0-1) - T_2(n_0)]. \tag{26}$$

Once again solving the above equation recursively leads to

$$T_2(n_0) = -\sum_{n=n_0+1}^{N} U(n), \tag{27}$$

where $U(n)$ is given by

$$U(n) = -\frac{2}{p_+} \sum_{i=0}^{n-1} K^{n-i-1} T(n). \tag{28}$$

For $n_0 = 0$, the second moment becomes

$$\langle T^2 \rangle = \frac{(1-K+6K^{N+1})}{p_+^2 (1-K)^3} \left[N + \frac{(1+K)}{(1-K+6K^{N+1})} N^2 - \frac{2K(1-K^N)(2+K^{N+1})}{(1-K)(1-K+6K^{N+1})} \right]. \tag{29}$$

Finally, the variance of the elongation time is given by

$$\sigma^2 = \langle T^2 \rangle - \langle T \rangle^2 = \frac{(1+K+K^{1+N})}{p_+^2 (1-K)^3} \left[N - \frac{K(1-K^N)(4+K+K^{1+N})}{(1-K)(1+K+4K^{1+N})} \right]. \tag{30}$$

In the limit $K \ll 1$ (polymerization is overwhelmingly favored over depolymerization) we can express the mean elongation time and variance up to first-order in K (see Eq. 2). In this regime, both the mean and the variance of the elongation time depend linearly on the template length N. Also the mean elongation time and variance approach the mean and variance

of the sum of N independent and identically distributed (i.i.d.) exponential steps. Since the sum of i.i.d. exponential random variables is γ-distributed we can assume that in the small K limit the elongation time, T, follows a γ-distribution

$$G(T|\alpha,\beta)=\frac{T^{\alpha-1}e^{-\frac{T}{\beta}}}{\Gamma(\alpha)\beta^{\alpha}}. \tag{31}$$

The parameters α and β can be calculated from the mean and variance using the relationships $\mu=\alpha\beta$ and $\sigma^2=\alpha\beta^2$:

$$\alpha=\frac{(N+KN-K)^2}{N+4KN-4K}, \tag{32a}$$

$$\beta=\frac{1}{p_+}\frac{N+4NK-4K}{N+NK-K}. \tag{32b}$$

In the limit of large N the distribution of elongation times approaches a Gaussian with mean and variance given by Eqs. 2a and 2b, respectively, in agreement with the Central Limit Theorem.

APPENDIX B: ELONGATION PAUSES AND BACKTRACKING

We model the dynamics of backtracking in terms of an unbiased random walk with rate c. For simplicity, we characterize backtracked states by $l=-m$ where $1\le l\le M$. The probability, $P(l,t)$, of finding the TEC in state l at time t given it was in state $l=1$ at $t=0$, follows the Master equation given in Eq. 3. By using the Laplace transform $\tilde{p}(l,s)=\int_0^\infty P(l,t)e^{-st}dt$, we can eliminate the time derivative in Eq. 3 and obtain an algebraic difference equation,

$$s\tilde{p}(l,s)-\delta_{l,1}=c\tilde{p}(l-1,s)+c\tilde{p}(l+1,s)-2c\tilde{p}(l,s), \tag{33}$$

where $\delta_{l,1}$ is the Kronecker delta.

Case 1: uninterrupted backtracking

In this case (see *schematic diagram* in Fig. 10 *a*), the boundary conditions for Eq. 3 are: $P(0,\ t)=0$ (absorbing) and $cP(M,\ t)=cP(M+1,\ t)$ (reflecting).

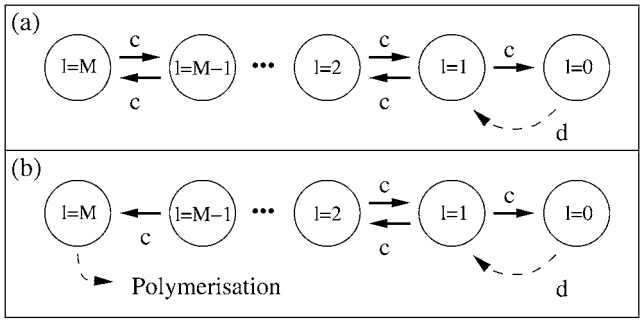


FIGURE 10 Schematic illustration of the two cases of restricted backtracking: (*a*) uninterrupted backtracking and (*b*) backtracking with transcript arrest. In both cases, variable l denotes the number of nucleotides that the TEC has translocated backward. Translocation is possible up to $l=M$. A backtracking pause commences with the TEC at state $l=1$ (*dashed arrow*) and terminates when state $l=0$ is reached. For the case of backtracking with transcript arrest, the TEC is halted at state $l=M$ and can resume polymerization only with the aid of accessory factors (*left dashed arrow*).

We solve Eq. 33 (as described in (39)), with boundary conditions $\tilde{p}(0,s)=0$ $c\tilde{p}(M,s)=c\tilde{p}(M+1,s)$, and obtain a closed formula for the Laplace transform of the probability flux to state $l=0$, $\tilde{F}(0,s)=c\tilde{p}(1,s)$,

$$\tilde{F}(0,s)=\frac{\sinh[M\phi(s)]-\sinh[(M-1)\phi(s)]}{\sinh[(M+1)\phi(s)]-\sinh[M\phi(s)]}, \tag{34}$$

where $\tanh\phi(s)=\sqrt{1-1/(s/2c+1)^2}$ The probability flux $F(0,\ t)$ is equivalent to the probability of exiting the pause at time t, and its Laplace transform, $\tilde{F}(0,s)$, evaluated at $s=0$, gives the probability of eventually exiting the pause (39). From Eq. 34, one obtains $\tilde{F}(0,s=0)=1$, i.e., the TEC will eventually exit the pause and resume elongation. $\tilde{F}(0,s)$ is also the moment-generating function containing all the positive integer moments of the exit time, as the coefficients of its power expansion in s (39). We expand Eq. 34 to get

$$\tilde{F}(0,s)=1-\frac{M}{c}s+\mathrm{O}(s^2), \tag{35}$$

from which we obtain the mean pause duration $\langle t\rangle=\mathrm{M/c}$.

We can also use $\tilde{F}(0,s)$ to calculate the distribution of pauses. In the limit $t\gg 1/\mathrm{c}$, i.e., for times much longer than the time for a single step, Eq. 34 becomes

$$\tilde{F}(0,s)\approx\frac{\cosh\left[\sqrt{\frac{s}{c}}(M)\right]}{\cosh\left[\sqrt{\frac{s}{c}}(M+1)\right]}. \tag{36}$$

By inverting the above Laplace transform (58), we can express the distribution of pause duration, $\mathcal{P}(t)\equiv F(0,t)$ (for times $>1/c$) in terms of the Jacobi θ_1 function,

$$\mathcal{P}(t)=a^{-1}\frac{\partial}{\partial\nu}\theta_1\left(\frac{1}{2}\nu a^{-1}\middle|ta^{-2}\right), \tag{37}$$

where $\nu=M/\sqrt{c}$, $a=(M+1)/\sqrt{c}$ and $\theta_1(z|t)$ can be expressed as the infinite series (58)

$$\theta_1(z|t)=\frac{1}{\sqrt{\pi t}}\sum_{\mathrm{n}=-\infty}^{\infty}(-1)^{\mathrm{n}}\exp[-(z+n-1/2)^2/t]. \tag{38}$$

Equation 37 leads to an expression for $\mathcal{P}(t)$. In particular, we obtain

$$\mathcal{P}(t)=\frac{-(M+1)}{\sqrt{\pi}\sqrt{c}t^{3/2}}\sum_{\mathrm{n}=-\infty}^{+\infty}(-1)^{\mathrm{n}}\exp\left[e^{-\frac{(1+M)^2}{ct}\left(n-\frac{1}{2(M+1)}\right)^2}\right]\left(n-\frac{1}{2(M+1)}\right). \tag{39}$$

Simpler expressions for $\mathcal{P}(t)$ can be obtained in the limits $t\ll M^2/c$ and $t\gg M^2/c$ (see Eq. 4 in main text). Plots of the analytic expression for $\mathcal{P}(t)$ along with the two asymptotic limits are shown in Fig. 11 *a*.

Case 2: backtracking with transcript arrest

In this case (see *schematic diagram* in Fig. 10 *b*) the boundary conditions imposed on Eq. 3 are: $P(0,\ t)=P(M,\ t)=0$. Once again, we solve Eq. 33 with boundary conditions $\tilde{p}(0,s)=\tilde{p}(M,s)=0$ to obtain a closed expression for the Laplace transforms of the exit probabilities to either boundary,

$$\tilde{F}(0,s)=\frac{\sinh[(M-1)\phi(s)]}{\sinh[M\phi(s)]}, \tag{40a}$$

$$\tilde{F}(M,s)=\frac{\sinh[\phi(s)]}{\sinh[M\phi(s)]}, \tag{40b}$$

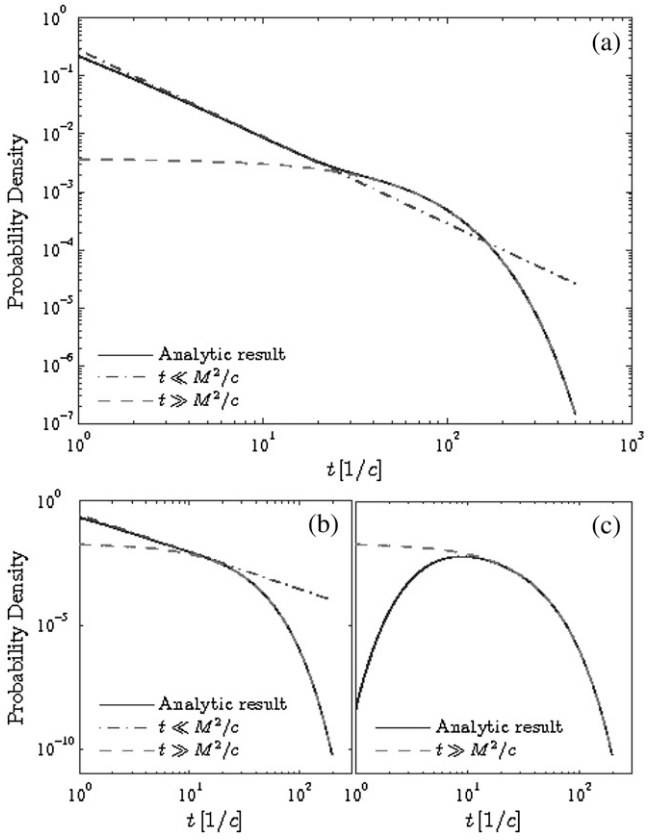


FIGURE 11 Analytic results for the duration of backtracking pauses, cases 1 and 2, for $M = 10$. (*a*) Case 1: restricted, uninterrupted backtracking. Probability distribution $\mathcal{P}(t)$ of exit time to absorbing boundary $l = 0$ in the presence of a reflecting boundary at $l = M$. Solid line corresponds to the analytic result Eq. 39, and dashed and dash-dotted lines to the two asymptotic limits in Eq. 4. (*b*, *c*). Case 2: restricted backtracking with transcript arrest. (*b*) Probability distribution $\mathcal{P}_0(t)$, of exit time to absorbing boundary $l = 0$ in the presence of an absorbing boundary at $l = M$. Solid line corresponds to the analytic result Eq. 45a, and dashed and dash-dotted lines to the two asymptotic limits in Eq. 5. (*c*) Probability distribution $\mathcal{P}_M(t)$ of exit time to absorbing boundary $l = M$ in the presence of an absorbing boundary at $l = 0$. Solid line corresponds to the analytic result Eq. 45b, and dashed line to the asymptotic limit in Eq. 5. In all cases, the initial state is assumed to be $l = 1$.

where $\tanh\phi(s) = \sqrt{1 - 1/(s/2c+1)^2}$. Evaluating the Laplace transforms at $s = 0$, we find that the TEC will eventually exit the pause either through state $l = M$ with probability $1/M$ or through state $l = 0$ with probability $1 - 1/M$. Once again, since Eq. 40a and Eq. 40b are generating functions, we can expand them in power series in s to obtain the mean exit times to either boundary, $\langle t\rangle_0$ and $\langle t\rangle_M$:

$$\langle t\rangle_0 = \frac{2M-1}{6c}, \tag{41a}$$

$$\langle t\rangle_M = \frac{M^2-1}{6c}. \tag{41b}$$

In the presence of accessory factors the arrested transcript is cleaved and the TEC returns to a polymerization competent state. If we assume that the accessory factors act on relatively fast timescales, then the overall mean pause duration is just the weighted sum of $\langle t\rangle_0$ and $\langle t\rangle_M$, $\langle t\rangle = (M-1)/2c$.

We can also use $\tilde{F}(0,s)$ and $\tilde{F}(M,s)$ to calculate the full distribution for the exit times to either boundary. For times much longer than the time for a single step, $t \gg 1/c$, Eqs. 40a and 40b become

$$\tilde{F}(0,s) \approx \frac{\sinh\left[\sqrt{\frac{s}{c}}(M-1)\right]}{\sinh\left[\sqrt{\frac{s}{c}}M\right]}, \tag{42a}$$

$$\tilde{F}(M,s) \approx \frac{\sinh\left[\sqrt{\frac{s}{c}}\right]}{\sinh\left[\sqrt{\frac{s}{c}}M\right]}. \tag{42b}$$

By inverting the above Laplace transforms (58), the distribution of exit times to the boundaries at $l = 0$, $\mathcal{P}_0(t) \equiv F(0,t)$, and at $l = M$, $\mathcal{P}_M(t) \equiv F(M,t)$ (for times much greater than $1/c$) can be expressed in terms of the Jacobi θ_4 function

$$\mathcal{P}_0(t) = a_0^{-1}\frac{\partial}{\partial\nu_0}\theta_4\left(\frac{1}{2}\nu_0 a_0^{-1}\middle| ta_0^{-2}\right), \tag{43a}$$

$$\mathcal{P}_M(t) = a_M^{-1}\frac{\partial}{\partial\nu_M}\theta_4\left(\frac{1}{2}\nu_M a_M^{-1}\middle| ta_M^{-2}\right), \tag{43b}$$

where $\nu_0 = (M-1)/\sqrt{c}$, $\nu_M = 1/\sqrt{c}$, $a_0 = a_M = M/\sqrt{c}$, and $\theta_4(z|t)$ can be expressed as the infinite series (58)

$$\theta_4(z|t) = \frac{1}{\sqrt{\pi t}}\sum_{n=-\infty}^{\infty}(-1)^n\exp[-(z+n+1/2)^2/t]. \tag{44}$$

Equations 43a and 43b lead to the following expressions for $\mathcal{P}_0(t)$ and $\mathcal{P}_M(t)$:

$$\mathcal{P}_0(t) = \frac{-M}{\sqrt{\pi}\sqrt{c}t^{3/2}}\sum_{n=-\infty}^{+\infty}\exp\left[e^{-\frac{M^2}{ct}\left(n-\frac{1}{2M}\right)^2}\right]\left(n-\frac{1}{2M}\right), \tag{45a}$$

$$\mathcal{P}_M(t) = \frac{-M}{\sqrt{\pi}\sqrt{c}t^{3/2}}\sum_{n=-\infty}^{+\infty}\exp\left[e^{-\frac{M^2}{ct}\left(n+\frac{M+1}{2M}\right)^2}\right]\left(n+\frac{M+1}{2M}\right). \tag{45b}$$

Simpler expressions for both $\mathcal{P}_0(t)$ and $\mathcal{P}_M(t)$ can be obtained in the limits $t \ll M^2/c$ and $t \gg M^2/c$ (see Eq. 5 in main text). Plots of the analytic expression for $\mathcal{P}_0(t)$ and $\mathcal{P}_M(t)$, along with the corresponding asymptotic limits are shown in Fig. 11, panels *b* and *c*, respectively.

APPENDIX C: TRANSCRIPTION WITH RESTRICTED BACKTRACKING

Parameter d, the transition rate from translocation state $m = 0$ to $m = -1$ (see Fig. 2), determines the density of backtracking. If we assume rapid transition between the active transition states $m = 0$ and $m = 1$, then at each template position the TEC can 1), proceed with polymerization, with rate $p_+ = k_f(b/(a+b))$; 2), proceed with depolymerization, with rate $p_- = k_b(a/(a+b))$; or 3), enter state $m = -1$, with an effective rate $d' = d(a/(a+b))$ (see Fig. 2 *b*). Therefore, at a given position n, the TEC enters a pause with probability

$$P_{PAUSE} = \frac{d'}{d' + p_+ + p_-}. \tag{46}$$

Since we assume that a pause occurs independently at each template position, we can estimate the probability P_{PAUSE} as the ratio of the expected number of pauses to the DNA template length i.e., $\delta/N = P_{PAUSE}$.

Simulations

Simulated data were generated using standard Monte Carlo techniques (Gillespie algorithm) (59,60), implemented in ANSI-C. At each step a random, exponentially distributed, number was generated that was used as the time interval until the next transition. The parameter, λ, of the exponential distribution was set equal to the sum of the transition rates to all accessible states. To decide to which state the transition will occur, a state was picked randomly from all accessible states with a probability proportional to the corresponding transition rate. The total elapsed time and the state were updated accordingly and the process was repeated.

In the case of Model A and for each set of parameter values, data were generated by 10^3 independent simulation runs. Since the values of parameters a and b are not known, arbitrary ones where used, which preserved the ratio found in the literature (see Table 1 of main text) and were higher than the rates of polymerization/depolymerization. In the case of the models of backtracking pauses and Model B, 10^5 simulations were performed for each set of parameter values to accurately capture the shape of the distribution and the scaling behavior. The parameters for Model B were selected so as to yield the experimentally observed values (19,22). In particular, a, b, k_f, and k_b were selected to yield an average velocity of 10 bp/s, while d was chosen to yield 1 and 10 pauses/kb. For simulations of the integrated initiation/elongation/degradation model the parameters used were selected to match the ones observed in Golding et al. (11): $N = 4$ kb, $L = 100$ bp, $M = 10$ bp, $p_+ = 50\ s^{-1}$, $K = 0.01$, $c = 0.1\ s^{-1}$, $k_i = 0.02\ s^{-1}$, and $k_d = 310^{-4}\ s^{-1}$ and $d' = 0.05\ s^{-1}$ (yielding 1 pause/kb).

Note added in proof: After submission we became aware of the recent experimental work by Galburt et al. (57), which studies the distribution of durations of pauses of RNAP II and finds a $t^{-3/2}$ dependence as predicted by Eqs. 4 and 5.

This work was supported by the Engineering and Physical Sciences Research Council under grants No. EP/D003105 and EP/C011953/1 (N.C.), the Medical Research Council under grant No. G0300556 (N.C., C.M.P., T.B.L.), the Royal Society (T.B.L.), and the University of Leeds (M.V.).

REFERENCES

1. Schrödinger, E. 1944. What is Life? Cambridge University Press, New York.
2. Paulsson, J. 2004. Summing up the noise in gene networks. *Nature*. 427:415–418.
3. Paulsson, J. 2005. Models of stochastic gene expression. *Phys. Life Rev.* 2:157–175.
4. Kærn, M., T. C. Elston, W. J. Blake, and J. J. Collins. 2005. Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Gen.* 6:451–464.
5. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain. 2002. Stochastic gene expression in a single cell. *Science*. 297:183–186.
6. Pedraza, J. M., and A. van Oudenaarden. 2005. Noise propagation in gene networks. *Science*. 307:1965–1969.
7. Yu, J., J. Xiao, X. Ren, K. Lao, and X. S. Xie. 2006. Probing gene expression in live cells, one protein molecule at a time. *Science*. 311: 1600–1603.
8. Greive, S. J., and P. H. von Hippel. 2005. Thinking quantitatively about transcriptional regulation. *Nat. Rev. Mol. Cell Biol.* 6:221–232.
9. von Hippel, P. H., and T. D. Yager. 1991. Transcription elongation and termination are competitive kinetic processes. *Proc. Natl. Acad. Sci. USA*. 88:2307–2311.
10. Swain, P. S., and A. Longtin. 2006. Noise in genetic and neural networks. *Chaos*. 16:026101.
11. Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox. 2005. Real-time kinetics of gene activity in individual bacteria. *Cell*. 123:1025–1036.
12. Bai, L., T. Santangelo, and M. D. Wang. 2006. Single-molecule analysis of RNA polymerase transcription. *Annu. Rev. Biophys. Biomol. Struct*. 35:343–360.
13. Braslavsky, I., B. Hebert, E. Kartalov, and S. R. Quake. 2003. Sequence information can be obtained from single DNA molecules. *Proc. Natl. Acad. Sci. USA*. 100:3960–3964.
14. Malan, T. P., A. Kolb, H. Buc, and W. R. McClure. 1984. Mechanism of CRP-cAMP activation of *lac* operon transcription initiation activation of the P1 promoter. *J. Mol. Biol.* 180:881–909.
15. McClure, W. R. 1985. Mechanisms and control of transcription initiation in prokaryotes. *Annu. Rev. Biochem.* 54:171–204.
16. Zhang, X., T. Reeder, and R. Schleif. 1996. Transcription activation parameters at ara pBAD. *J. Mol. Biol.* 258:14–28.
17. Skinner, G. M., C. G. B. D. M. Quinn, J. E. Molloy, and J. G. Hoggett. 2004. Promoter binding, initiation, and elongation by bacteriophage T7 RNA polymerase. *J. Biol. Chem.* 279:3239–3244.
18. Saunders, A., L. J. Core, and J. T. Lis. 2006. Breaking barriers to transcription elongation. *Nat. Rev. Mol. Cell Biol.* 7:557–567.
19. Shaevitz, J. W., E. A. Abbondanzieri, R. Landick, and S. M. Block. 2003. Backtracking by single RNA polymerase molecules observed at near base pair resolution. *Nature*. 426:684–687.
20. Neuman, K. C., E. A. Abbondanzieri, R. Landick, J. Gelles, and S. M. Block. 2003. Ubiquitous transcriptional pausing is independent of RNA polymerase backtracking. *Cell*. 115:437–447.
21. Herbert, K. M., A. L. Porta, B. J. Wong, R. A. Mooney, K. C. Neuman, R. Landick, and S. Block. 2006. Sequence-resolved detection of pausing by single RNA polymerase molecules. *Cell*. 125:1083–1094.
22. Forde, N. R., D. Izhaky, G. R. Woodcock, G. J. L. Wuite, and C. Bustamante. 2002. Using mechanical force to probe the mechanism of pausing and arrest during continuous elongation by *Escherichia coli* RNA polymerase. *Proc. Natl. Acad. Sci. USA*. 99:11682–11687.
23. Zenkin, N., Y. Yuzenkova, and K. Severinov. 2006. Transcript-assisted transcriptional proofreading. *Science*. 313:518–520.
24. Nickels, B. E., and A. Hochschild. 2004. Regulation of RNA polymerase through the secondary channel. *Cell*. 118:281–284.
25. Mote, J. J., and D. Reines. 1998. Recognition of a human arrest site is conserved between RNA polymerase II and prokaryotic RNA polymerases. *J. Biol. Chem.* 273:16843–16852.
26. Guajardo, R., and R. Sousa. 1997. A model for the mechanism of polymerase translocation. *J. Mol. Biol.* 256:8–19.
27. Julicher, F., and R. Bruinsma. 1998. Motion of RNA polymerase along DNA: a stochastic model. *Biophys. J.* 74:1169–1185.
28. Bai, L., A. Shundrovsky, and M. D. Wang. 2004. Sequence-dependent kinetic model for transcription elongation by RNA polymerase. *J. Mol. Biol.* 344:335–349.
29. Tadigotla, V. R., D. O. Maoiléidigh, A. M. Sengupta, V. Epshtein, R. H. Ebright, E. Nudler, and A. E. Ruckenstein. 2006. Thermodynamic and kinetic modeling of transcription pausing. *Proc. Natl. Acad. Sci. USA*. 103:4439–4444.
30. Raj, A., C. S. Peskin, D. Tranchina, D. Y. Vargas, and S. Tyagi. 2006. Stochastic mRNA synthesis in mammalian cells. *PLOS Biol.* 4:1707–1719.
31. Chubb, J. R., T. Trcek, S. M. Shenoy, and R. H. Singer. 2006. Transcriptional pulsing of a developmental gene. *Curr. Biol.* 16:1018–1025.
32. Korzheva, N., A. Mustaev, M. Kozlov, A. Malhorta, V. Nikiforov, A. Goldfarb, and S. A. Darst. 2000. A structural model of transcription elongation. *Science*. 289:619–625.
33. Nudler, E., A. Mustaev, 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.
34. 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*. 279:1755–1760.
35. Borukhov, S., V. Sagitov, and A. Goldfarb. 1993. Transcript cleavage factor from *E. coli*. *Cell*. 72:459–466.

36. Fish, R. N., and C. M. Kane. 2002. Promoting elongation with transcript cleavage stimulatory factors. *Biochim. Biophys. Acta.* 1577:287–307.

37. van Kampen, N. G. 1992. Stochastic Processes in Physics and Chemistry. Elsevier, New York.

38. Gardiner, C. W. 2004. Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences, 3rd Ed. Springer-Verlag, Berlin, Germany.

39. Redner, S. 2001. A Guide to First-Passage Processes. Cambridge University Press, New York.

40. Gotta, S. L., O. L. Miller, and S. L. French. 1991. rRNA transcription rate in *Escherichia coli*. *J. Bacteriol.* 173:6647–6649.

41. Marko, J. F., and E. D. Siggia. 1995. Statistical mechanics of supercoiled DNA. *Phys. Rev. E.* 52:2912–2938.

42. Col, A. D., and T. B. Liverpool. 2004. Statistical mechanics of double-helical polymers. *Phys. Rev. E.* 69:61907–61911.

43. Derrida, B., S. A. Janowsky, J. L. Lebowitz, and E. R. Speer. 1993. Exact solution of the totally asymmetric simple exclusion process: shock profiles. *J. Stat. Phys.* 73:813–842.

44. Evans, M. R., D. P. Foster, C. Godrèche, and D. Mukamel. 1995. Spontaneous symmetry breaking in a one-dimensional driven diffusive system. *Phys. Rev. Lett.* 74:208–211.

45. Bremer, H., and M. Ehrenberg. 1995. Guanosine tetraphosphate as a global regulator of bacterial RNA synthesis: a model involving RNA polymerase pausing and queuing. *Biochim. Biophys. Acta.* 1262: 15–36.

46. Artsimovitch, I., and R. Landick. 2000. Pausing by bacterial RNA polymerase is mediated by mechanistically distinct classes of signals. *Proc. Natl. Acad. Sci. USA.* 97:7090–7095.

47. Li, B., M. Carey, and J. L. Workman. 2007. The role of chromatin during transcription. *Cell.* 128:707–719.

48. Carey, M., B. Li, and J. L. Workman. 2006. RSC Exploits histone acetylation to abrogate the nucleosomal block to RNA polymerase II elongation. *Mol. Cell.* 24:481–487.

49. van Zon, J. S., M. J. Morelli, S. T. Tanase, and P. R. ten Wolde. 2006. Diffusion of transcription factors can drastically enhance the noise in gene expression. *Biophys. J.* 91:4350–4367.

50. Blank, A., J. A. Gallant, R. R. Burgess, and L. A. Loeb. 1986. An RNA polymerase mutant with reduced accuracy of chain elongation. *Biochemistry.* 25:5920–5928.

51. Reference deleted in proof.

52. Rasmussen, E. B., and J. T. Lis. 1993. *In vivo* transcriptional pausing and cap formation on three *Drosophila* heat shock genes. *Proc. Natl. Acad. Sci. USA.* 90:7923–7927.

53. Adelman, K., M. T. Marr, J. Werner, A. Saunders, Z. Ni, E. D. Andrulis, and J. T. Lis. 2005. Efficient release from promoter-proximal stall sites requires transcript cleavage factor TFIIS. *Mol. Cell.* 17: 103–112.

54. Lee, R. C., and V. Ambros. 2001. An extensive class of small RNAs in *Caenorhabditis elegans* complementarity to lin-14. *Science.* 295:862–864.

55. Lagos-Quintana, M., R. Rauhut, W. Lendeckel, and T. Tuschl. 2001. Identification of novel genes coding for small expressed RNAs. *Science.* 295:853–858.

56. Lau, N. C., L. P. Lim, E. G. Weinstein, and D. P. Bartel. 2001. An abundant class of tiny RNAs with probable regulatory roles in *Caenorhabditis elegans*. *Science*. 295:858–862.

57. Galburt, E. A., S. W. Grill, A. Wiedmann, L. Lubkowska, J. Choy, E. Nogales, M. Kashlev, and C. Bustamante. 2007. Backtracking determines the force sensitivity of RNAP II in a factor-dependent manner. *Nature*. 446:820–823.

58. Oberhettinger, F., and L. Badii. 1973. Tables of Laplace Transforms. Springer-Verlag, Berlin, Germany.

59. Gillespie, D. T. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–2361.

60. Binder, K. 1986. Monte Carlo Methods in Statistical Physics, 2nd Ed. Springer-Verlag, Berlin, Germany.

61. Kingston, R. E., W. C. Nierman, and M. J. Chamberlin. 1981. A direct effect of guanosine tetraphosphate on pausing of *Escherichia coli* RNA polymerase during RNA chain elongation. *J. Biol. Chem.* 256:2787–2797.

62. Yin, H., M. D. Wang, K. Svodoba, R. Landick, and S. M. Block. 1995. Transcription against an applied force. *Science*. 270:1653–1657.

63. Schafer, D. A., J. Gelles, M. P. Sheetz, and R. Landick. 1991. Transcription by single molecules of RNA polymerase observed by light microscopy. *Nature*. 352:444–448.