

Fixed and distributed delays from a family of gene transcription models

S. Hossein Hosseini* and Marc R. Roussel†

Alberta RNA Research and Training Institute, Department of Chemistry and Biochemistry, University of Lethbridge,
4401 University Dr W, Lethbridge, AB T1K 3M4, Canada

(Dated: February 14, 2019)

Models intended to describe the time evolution of a gene network must somehow include transcription, the DNA-templated synthesis of RNA, and translation, the RNA-templated synthesis of proteins. In eukaryotes, the DNA template for transcription can be very long, often consisting of tens of thousands of nucleotides, and lengthy pauses may punctuate this process. Accordingly, transcription can last for many minutes, in some cases hours. There is a long history of introducing delays in gene expression models to take the transcription and translation times into account. Here we study a family of detailed transcription models that includes initiation, elongation, and termination reactions. We establish a framework for computing the distribution of transcription times, and work out these distributions for some typical cases. For elongation, we find that a simple fixed delay is a good model provided there are no sites where long pauses occur. The initiation and termination phases of the model then generate a nontrivial delay distribution, and elongation shifts this distribution by an amount corresponding to the elongation delay. If there are long pauses during elongation, because of the modularity of the family of models considered, the elongation phase can be partitioned into reactions generating a simple delay (elongation through regions where there are no long pauses), and reactions whose distribution of waiting times must be considered explicitly (initiation, termination, and motion through regions where long pauses are likely). In these cases, the distribution of transcription times again involves a nontrivial part and a shift due to normal elongation processes.

PACS numbers: 87.14.G-, 87.10.Mn, 82.39.Pj

I. INTRODUCTION

Transcription is a complex process for the biosynthesis of RNA process from a DNA template. Roughly, this process can be divided into three phases: initiation, elongation and termination [1]. During initiation, a set of protein factors bind the DNA template upstream of the start site and facilitate the binding and positioning of RNA polymerase [2, 3]. The first few nucleotides are added to the nascent RNA, which eventually transitions into productive elongation. Elongation consists of a sequence of rapid nucleotide addition steps [4]. While each individual step is fast (over 70 nt/s [4]; nt = nucleotide), genes can be very long. The average human gene, which contains many introns that must be spliced out to form the final transcript, has a length of 28 000 nt [5], and much longer genes are known [6], so the elongation time can be significant. Finally, transcription must be terminated. In eukaryotes, redundant mechanisms cooperate to ensure termination [7].

Transcription is only the first step in gene expression. In eukaryotes, the primary transcript is subjected to processing, which includes splicing, 5' capping, and 3' polyadenylation; the completed messenger RNA (mRNA) is packaged with proteins into a messenger ribonucleoprotein (mRNP) complex that is required for

export from the nucleus; it must make its way to the nuclear pore and be exported to the cytoplasm; it may need to be shuttled to a specific part of the cell; and finally it is translated to a protein. Since splicing [8], 5' capping [9] and mRNP assembly [10] are generally co-transcriptional, a minimal model of gene expression in eukaryotes might consider just transcription, nuclear export, and translation. (Polyadenylation is strongly coupled to transcription termination, so it could be included in the termination phase of a transcription model.) All three of these major processes may be separately regulated, so it is not, in general, possible to collapse them into a single protein-synthesis step as is often done in gene expression models. If we want to model a gene network however, we almost certainly do not want to model transcription, nuclear export and translation in detail. Rather, we would prefer to treat each of these processes as effective reactions; one reaction for transcription, one for export, and one for translation. If we intend to model the temporal evolution of a gene network, it may be necessary to consider the time required for each of these processes. One way to avoid detailed models of these processes while still considering the expression time of a gene is to incorporate delays, which may be fixed or distributed, into a model containing effective reactions. But this of course means that we need to know something about the distributions of these delays.

In this paper, we will compute the transcriptional delay distribution, the transcriptional delay being defined here as the delay before the appearance of the product RNA *given that* the gene is active at time zero, in a class of models of varying degrees of complexity. Although the

*Electronic address: hosseini@uleth.ca

†Electronic address: roussel@uleth.ca; URL: <http://people.uleth.ca/~roussel>

transcription model studied here is similar to prokaryotic models presented elsewhere [11, 12], we focus here on a model and on parameters appropriate to eukaryotes [13]. We further narrow the focus to protein-coding genes transcribed by RNA polymerase II (RNAPII). In eukaryotes, most protein-coding genes are rarely transcribed, so that it will be rare for two polymerases to interact [14]. Accordingly, we study a single-polymerase model.

There have been a number of kinetic models regarding the kinetics of transcription. Several models have focused on the mechanics and statistical thermodynamics of elongation [15–19], and some of these models have yielded insights into the factors responsible for the transcriptional delay (also known as a dwell time) [19].

In prior work, we proposed a single-gene transcription model from which we obtained analytical and numerical results for the distribution of the transcriptional delay [11]. We also studied the case where many polymerases transcribe the same gene and obtained stochastic simulation results [11–13]. A similar approach was followed by von Hippel’s group, obtaining time-dependent solutions from which a delay distribution could be inferred [20]. Of course, there is a formal similarity between transcription and translation, which are both templated processes, so we find similar work also in translation modeling. Garai and coworkers have computed both dwell times at a codon, and distributions of translation times [21, 22]. Perhaps most closely related to the present contribution is the work of Mier-y-Terán-Romero and coworkers, who showed that for a long mRNA, an advection equation is obtained in the continuum limit such a fixed elongation delay emerges [23].

Single-molecule optical trapping experiments show that the movement of RNA polymerase along the DNA strand is not continuous and is punctuated many times by pauses [24, 25]. Mechanistically, there are two main classes of pauses: (1) backtrack pausing, in which RNAPII slides backward reversibly along both the DNA and the RNA, (2) non-backtrack pausing, in which conformational changes in the RNA polymerase active site stop the nucleotide addition cycle [24, 26]. Backtracking pauses can be affected by the presence of a trailing RNA polymerase that can restrict how far an RNA polymerase can backtrack [27], leading to density dependence of the transcription rate [28]. Voliotis *et al.* [29] studied a model of backtracking transcriptional pauses, for which they obtained a distribution of transcription times (delays). They found that backtracking leads to a heavy-tailed distribution, a conclusion also reached by Klumpp [28].

Pauses can occur in many parts of a gene, and for many different reasons. In eukaryotes, the polymerase almost always pauses just downstream of the start site in an event known as promoter-proximal pausing [30–32]. These pauses are thought to have a regulatory role. Pausing can be associated with features of the sequence [18, 33]. Recent work has shown that splicing also causes pausing [34]. Pausing also facilitates termination of tran-

scription [35].

In this paper, we treat transcription both with and without pauses, and we consider both short ubiquitous pauses and specific pause sites where lengthy pauses occur. We first present a simplified version of Vashishta’s model of eukaryotic transcription [13]. Then we utilize this model to write the corresponding chemical master equations for each phase. For both initiation and termination phases, since there are few ordinary differential equations (ODEs), the probability functions for being in each state of the model can be easily obtained. For the elongation phase, we again write down the chemical master equation. However, as a large number of ODEs is required to describe this phase, we derive a continuum limit of these ODEs and find the corresponding Fokker-Planck equation. Our approach here is similar in spirit, if not in detail, to the work of Mier-y-Terán-Romero and coworkers on translation [23]. In sufficiently long genes, the effect of the elongation phase is to introduce a fixed delay between initiation and termination. It is then possible to obtain an analytic expression for the distribution of the total transcription time. We then go on to consider pauses. Short ubiquitous pauses are shown to increase the elongation delay but have no other effect on the shape of the distribution of transcription times. Because of the modularity of the class of models considered [20] such that the total delay is just a sum of delays at each nucleotide, we can treat the pause site(s) separately, and then get the distribution of the total transcriptional delay as a convolution.

II. EUKARYOTIC TRANSCRIPTION MODEL

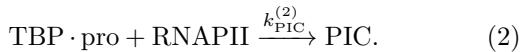
We start with a simplified version of Vashishta’s kinetic model for the eukaryotic transcription process [13], removing abortive initiation, early elongation/promoter escape and promoter-proximal pausing. Promoter-proximal pausing can, at least to a first approximation, be modeled as a distinct, long pause, as we will do in section IV. Similarly the early elongation phase could be treated analogously to productive elongation, but with different kinetic parameters. Abortive initiation is now thought to be much less significant in eukaryotes than in prokaryotes due to mechanical differences in initiation [36]. We leave the addition of these refinements to later work.

The model is divided into the usual three phases: initiation, elongation, and termination. This section is correspondingly organized.

A. Initiation

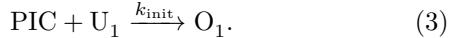
For the initiation phase, first the initiation factors, including the TATA-binding protein (TBP), bind to the TATA box of the promoter region (pro) of the DNA [reaction (1)], and then RNAPII locates the promoter and

binds to it, forming the pre-initiation complex (PIC) [reaction (2)] [2]:



The rate constants $k_{\text{bind}}^{(2)}$ and $k_{\text{PIC}}^{(2)}$ are second-order rate constants; the superscripts appear for emphasis.

After forming a stable PIC, the RNA polymerase moves to the first nucleotide of the DNA template. Reaction (3) represents the translocation of the active site of the RNA polymerase to the first nucleotide (nucleotide 1) if this nucleotide is unoccupied (U state), thus converting U_1 into an occupied nucleotide (O state):



Since we only consider a single polymerase here, the forward site will always be unoccupied. We maintain this notation for consistency with our earlier work [11–13].

From here and until we reach the termination phase, transcription consists of cycling through binding and recognition of a nucleoside triphosphate complementary to the template nucleotide, translocation to the next template position, and formation of a phosphodiester bond, which adds the selected nucleotide to the growing RNA chain. The early steps are slower than the corresponding steps in productive elongation, but for the reasons explained above, we ignore these kinetic differences in this contribution. Thus, the binding of the first nucleotide is treated as indistinguishable from binding of any other nucleotide and is included in the elongation part of the model.

B. Elongation

Although conceptually there are at least three distinct processes in elongation (binding of a complementary nucleoside triphosphate, translocation, catalysis of phosphodiester bond formation), the simplest models of transcription describe the process in terms of two states [11, 37], roughly corresponding to the pre- and post-translocated states. More [18, 19, 38] or fewer [20] states can be used to model elongation. In our models [11–13], the two states are called occupied (O) and activated (A) (Fig. 1), corresponding respectively to the pre- and post-translocated states.

The process of activation and translocation of RNA polymerase repeats itself as RNAPII translocates nucleotide by nucleotide along the gene. These repeated reactions are responsible for productive elongation [reactions (4) and (5)].

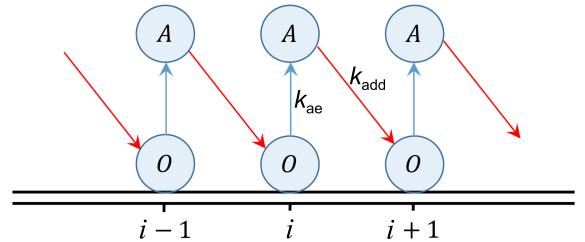
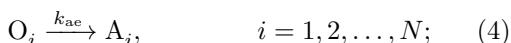
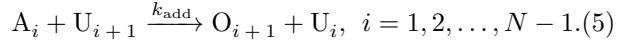
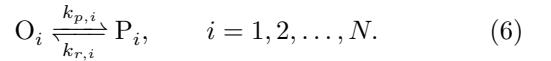


FIG. 1: Two internal states for each nucleotide during the elongation phase: occupied (O) and activated (A)



Here, N is the length of the template, i.e. the transcribed sequence, which is assumed fixed in the current model.

During elongation, the polymerase may pause for any of a number of reasons: There are so-called “ubiquitous” pauses which may occur in any part of the gene [25], promoter-proximal pauses associated with a change in state of the polymerase as it transitions from initiation to elongation [39], pauses associated with splicing [40], and pauses associated with termination [35]. Pausing models can be more or less complex and incorporate backtracking [29] or arrest [41]. Here we opt for a simple model of a reversible pause. We neglect backtracking, which is likely appropriate for short, ubiquitous pauses, but may be less so for long pauses. It appears that pausing proceeds from the pre-translocated (O) state [41, 42], so we write



We allow here for the possibility that the rate constants for pausing, $k_{p,i}$, and for the release from the pause, $k_{r,i}$, might be template-dependent. On the other hand, we need make no such allowance for the rate constants associated with productive elongation, k_{ae} and k_{add} , which are believed to be constant over a gene and across genes within a particular organism [43].

C. Termination

According to the allosteric model of termination, transcription of the poly(A) signal at the end of the gene can cause a termination-inducing conformational change in the elongation complex [44]. After the last reaction of elongation [activation at nucleotide N , reaction (4)], the elongation complex is converted into a termination complex (TC) [reaction (7)]. Finally, the termination complex dissociates and releases the RNA [reaction (8)].



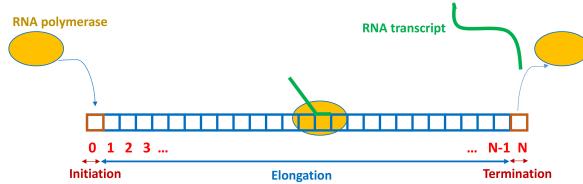


FIG. 2: A schematic representation of RNA polymerase on the DNA template strand.

The above reactions assume 100% efficiency of allosteric termination. In reality, there is at least one, and possibly two failsafe mechanisms backing up allosteric termination [7]. We assume a single mechanism here for simplicity. In terms of the completion of the transcript, the torpedo model will behave similarly to allosteric termination since its first step is to cleave the transcript at the poly(A) sequence, releasing it from the transcriptional machinery [7].

III. TIME-DEPENDENT PROBABILITY DISTRIBUTION FOR TRANSCRIPTION BY RNAPII

A. Initiation

We assume that all initiation processes are happening at $x = 0$ and all the termination processes are happening at $x = N$ (the nucleotide at the end of the transcribed sequence) (Fig. 2). By solving the chemical master equation (CME) for each state in the three phases of transcription, we can get time-dependent probability distributions for the progress of transcription, from which we can recover the transcriptional delay.

In principle, the propensities of reactions (1) and (2) depend on the populations of available TBP and RNAPII, respectively. Here, we assume that there is a constant cellular pool of these two species so that we can treat reactions (1) and (2) as pseudo-first order reactions, with rate constants $k_{\text{bind}} = k_{\text{bind}}^{(2)}[\text{TBP}]$, and $k_{\text{PIC}} = k_{\text{PIC}}^{(2)}[\text{RNAPII}]$, respectively. The governing master equations are

$$\frac{dP_{\text{pro}}(t)}{dt} = -k_{\text{bind}} P_{\text{pro}}(t), \quad (9)$$

$$\frac{dP_{\text{TBP}\cdot\text{pro}}(t)}{dt} = k_{\text{bind}} P_{\text{pro}}(t) - k_{\text{PIC}} P_{\text{TBP}\cdot\text{pro}}(t), \quad (10)$$

$$\frac{dP_{\text{PIC}}(t)}{dt} = k_{\text{PIC}} P_{\text{TBP}\cdot\text{pro}}(t) - k_{\text{init}} P_{\text{PIC}}(t), \quad (11)$$

with initial conditions $P_{\text{pro}}(0) = 1$, $P_{\text{TBP}\cdot\text{pro}}(0) = P_{\text{PIC}}(0) = P_{\text{O}_1}(0) = 0$. In these equations, $P_{\text{pro}}(t)$ is the probability that the promoter has never been bound up to time t , i.e. it is the survival probability of the empty promoter. Since we treat the single-polymerase case, we do not need to consider promoter clearance. $P_{\text{TBP}\cdot\text{pro}}(t)$ is the probability that, at time t , TBP

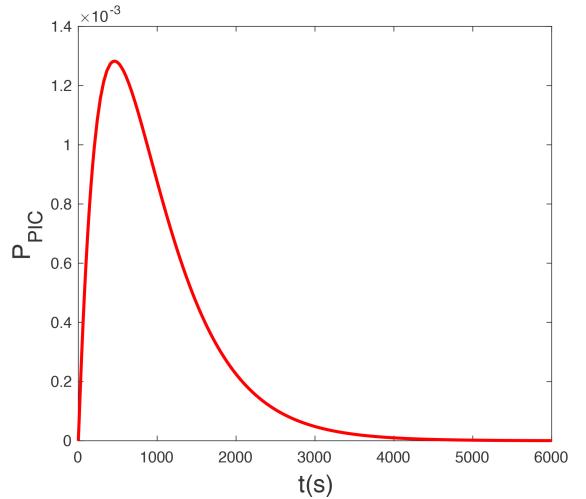


FIG. 3: $P_{\text{PIC}}(t)$ from Eq. 13 for $k_{\text{bind}} = 0.0016s^{-1}$, $k_{\text{PIC}} = 0.0029s^{-1}$ and $k_{\text{init}} = 0.6s^{-1}$.

is bound to the promoter. Similarly, $P_{\text{PIC}}(t)$ is the probability that a complete PIC is assembled at the promoter at time t . The time-dependent rate at which probability exits initiation and enters elongation is, from reaction (3),

$$v_{\text{init}} = k_{\text{init}} P_{\text{PIC}}. \quad (12)$$

The linear ordinary differential equations (9) to (11) can be solved to find the probability distribution of being in each state. The solution for $P_{\text{PIC}}(t)$ is

$$\begin{aligned} P_{\text{PIC}}(t) = & \frac{k_{\text{PIC}} k_{\text{bind}}}{(k_{\text{PIC}} - k_{\text{bind}})(k_{\text{bind}} - k_{\text{init}})(k_{\text{init}} - k_{\text{PIC}})} \\ & \times \left[(k_{\text{PIC}} - k_{\text{init}})e^{-k_{\text{bind}}t} + (k_{\text{init}} - k_{\text{bind}})e^{-k_{\text{PIC}}t} \right. \\ & \left. + (k_{\text{bind}} - k_{\text{PIC}})e^{-k_{\text{init}}t} \right]. \end{aligned} \quad (13)$$

Figure 3 shows $P_{\text{PIC}}(t)$. The parameters used here were estimated by Vashishtha [13]. Briefly, the rate of TBP binding to the TATA box is assumed not to be limited by reaction (1), but by dissociation of TBP dimers, so the effective rate constant for this non-elementary step is the rate constant for dimer dissociation, $k_{\text{bind}} = 0.0016s^{-1}$ [45]. The rate constant for PIC assembly, $k_{\text{PIC}} = 0.0029s^{-1}$, was taken from an *in vitro* study [46]. Finally, based on experimental data for the progress of the polymerase through the first few nucleotides [46], Vashishtha inferred a lower bound for k_{init} of $0.3s^{-1}$ [13]. The value of k_{init} eventually adopted is somewhat arbitrary, but in any event, it is clear that binding of TBP and PIC assembly are the rate limiting steps. Note that eukaryotic initiation times are very long, supporting our decision to focus on a single-polymerase model.

B. Elongation

1. Continuum approximation

The biggest challenge in dealing with elongation is having a large number of nucleotides, each of which must go through the elongation cycle consisting of reactions (4) and (5), with pausing as an added possibility. If we write the master equation for every state at all nucleotides, we will have a set of about $3N$ (N is number of nucleotides) ODEs which will be difficult to solve since, for typical protein-coding genes in eukaryotes, N might range from several hundred nucleotides to several tens of thousands [5]. However, the lengths of typical genes and the high rates of transcription [4] create an opportunity to make a continuum approximation.

Denote by $P_{i,\sigma}$ the probability that the polymerase has its active site at nucleotide i and that it is in state σ , where σ can take on the values O, P or A. The master equations governing the time evolution of the elongation phase are

$$\frac{dP_{i,O}(t)}{dt} = k_{\text{add}}P_{i-1,A}(t) - k_{\text{ae}}P_{i,O}(t) - k_{p,i}P_{i,O}(t) + k_{r,i}P_{i,P}(t), \quad (14)$$

$$\frac{dP_{i,A}(t)}{dt} = k_{\text{ae}}P_{i,O}(t) - k_{\text{add}}P_{i,A}(t), \quad (15)$$

$$\frac{dP_{i,P}(t)}{dt} = k_{p,i}P_{i,O}(t) - k_{r,i}P_{i,P}(t), \quad (16)$$

where $i = 1, 2, 3, \dots, N - 1$.

Because the states are mutually exclusive, the probability that the polymerase is at nucleotide i in any state, P_i , is the sum of the probabilities $P_{i,\sigma}$, that is

$$P_i(t) = P_{i,O}(t) + P_{i,A}(t) + P_{i,P}(t). \quad (17)$$

Using Eqs. (14)–(17), we obtain

$$\frac{dP_i}{dt} = k_{\text{add}}(P_{i-1,A} - P_{i,A}). \quad (18)$$

We want to treat the motion on the time scale of elongation. We therefore introduce the rescaled time

$$\theta = t/\tau_e, \quad (19)$$

where τ_e is the mean total elongation time. This quantity will be computed below. Introducing this scaling in Eq. (18), we get

$$\frac{dP_i}{d\theta} = k_{\text{add}}\tau_e(P_{i-1,A} - P_{i,A}). \quad (20)$$

For long genes, an increment of one nucleotide is a small change in position measured relative to the length of the gene. Therefore define

$$x = i/N, \quad (21)$$

$$\delta = 1/N. \quad (22)$$

In a continuous representation, the $P_{i,A}(\theta)$, which are supported on the natural numbers, will become continuous functions of x , $\rho_A(x, \theta)$, with the correspondence $P_{i,A}(\theta) = \delta\rho_A(i\delta, \theta)$. Similarly, we define the probability density for all states combined by $P_i(\theta) = \delta\rho_e(i\delta, \theta)$. Because δ is small, from the definition of the derivative, we have

$$\begin{aligned} P_{i-1,A}(\theta) - P_{i,A}(\theta) &= \delta(\rho_A(x - \delta, \theta) - \rho_A(x, \theta)) \\ &\approx -\delta^2 \frac{\partial \rho_A(x, \theta)}{\partial x}. \end{aligned} \quad (23)$$

This will hold provided the $P_{i,A}$ vary sufficiently slowly. Otherwise, the estimate of the derivative on which (23) is based is meaningless. This will fail if the gene contains a specific site at which pauses of long duration occur, where probability may accumulate for a time during elongation. The ultimate cause of this accumulation is a very rapid change in kinetic parameters over a small region of a gene. We return to this problem in section IV. In the meantime, we note that the treatment of translation by Mier-y-Terán-Romero and coworkers also required sufficiently slowly varying parameters.

Substituting (23) and the definitions of the densities into (20), we get the PDE

$$\frac{\partial \rho_e}{\partial \theta} = -k_{\text{add}}\tau_e\delta \frac{\partial \rho_A}{\partial x}. \quad (24)$$

Equation (24) is not closed as it relates the time derivative of ρ_e to the gradient of ρ_A . If we examine the state of a specified nucleotide from an ensemble of similarly prepared systems at an arbitrary moment in time, the conditional probability of catching it in a particular state given that the polymerase has reached this nucleotide is proportional to its mean residence time in that state. In particular

$$P_{i,A} = P_i P(A|i) = P_i \frac{\tau_{i,A}}{\tau_i}, \quad (25)$$

where $\tau_{i,A} = k_{\text{add}}^{-1}$ is the mean residence time in the A state at nucleotide i , and τ_i is the total residence time at nucleotide i . Converting to densities, Eq. (24) becomes

$$\frac{\partial \rho_e}{\partial \theta} = -\delta\tau_e \frac{\partial}{\partial x} \left(\frac{\rho_e}{\tau(x)} \right), \quad (26)$$

where $\tau(i\delta) = \tau_i$, and $\tau(x)$ interpolates between these values for $x \neq i\delta$. A similar equation has been obtained by Mier-y-Terán-Romero and coworkers for translation by the ribosome [23]. The approach used by these authors to solve their variable-velocity advection equation could be followed here.

Rather, we will focus for now on the case in which the kinetic parameters are identical at every nucleotide. Then, $\tau(x) = \tau_i$ is a constant, and

$$\tau_e = (N - 1)\tau_i + k_{\text{ae}}^{-1} \approx N\tau_i \quad (27)$$

for large N . Taking into account (22), (26) reduces to

$$\frac{\partial \rho_e}{\partial \theta} = -\frac{\partial \rho_e}{\partial x}, \quad (28)$$

i.e. a simple advective equation with unit velocity. (An alternative derivation of this equation is to be found in Ref. [47].) At the left boundary ($x = 0$), probability flows in at a rate $k_{\text{init}}P_{\text{PIC}}(t)$ [reaction (3)] or, in the scaling (19), $\tau_e k_{\text{init}}P_{\text{PIC}}(\tau_e\theta)$. The appropriate boundary condition for (28) is therefore

$$\rho_e(0, \theta) = \tau_e k_{\text{init}}P_{\text{PIC}}(\tau_e\theta). \quad (29)$$

(Technically, since the nucleotides are numbered from 1, the boundary condition should be applied at $x = \delta$ rather than $x = 0$. However, since δ is assumed small, there is no significant error in the boundary condition given above.) Since we deal with just one polymerase, and since this polymerase cannot start transcription before $t = 0$, we also have the initial condition $\rho_e(x, 0) = 0$.

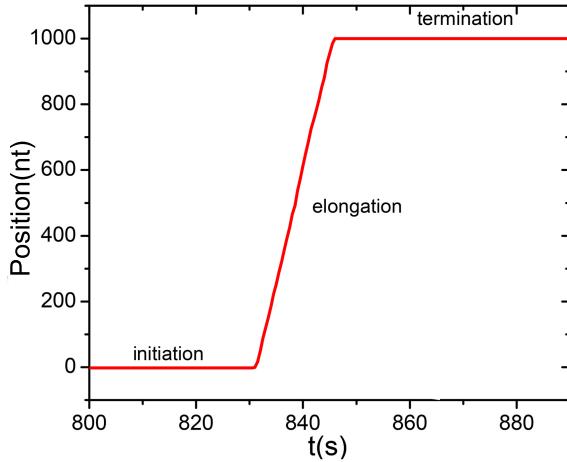


FIG. 4: Position of the RNA polymerase vs time obtained from the stochastic simulation of a single polymerase in the absence of pausing. Parameter values: $k_{ae} = 144 \text{ s}^{-1}$, $k_{add} = 144 \text{ s}^{-1}$, $k_{PIC} = 0.0029 \text{ s}^{-1}$, $k_{bind} = 0.0016 \text{ s}^{-1}$ and $k_{init} = 0.6 \text{ s}^{-1}$, $N = 1000$.

The PDE (28) with the boundary condition (29) simply transports probability through the elongation compartment at unit velocity. Accordingly, the solution is

$$\rho_e(x, \theta) = \tau_e k_{init} P_{PIC}(\tau_e(\theta - x)) H(\theta - x), \quad (30)$$

where $H(\cdot)$ is the Heaviside function. The flux exiting elongation, at $x = 1$, transformed back to the original time scale, is therefore

$$J_e = k_{init} P_{PIC}(t - \tau_e) H(t - \tau_e). \quad (31)$$

This solution holds for any fixed kinetic parameters, with or without pausing. We conclude that for transcription, as was the case for translation [23], elongation acts as a simple delay.

Figure 4 shows the position of the polymerase against time obtained from the stochastic simulation for a single polymerase in the absence of pausing, while Fig. 5 shows a similar simulation result with frequent short pauses. These simulations confirm that when there are no long pauses along a large gene, the elongation process is advective and RNAPII moves with a roughly constant velocity. Interestingly, the relatively constant time required to get through elongation has been observed experimentally [1].

2. The elongation delay

Equation (31) shows that, at least for fixed kinetic parameters, elongation only contributes a delay. Mier-y-Terán-Romero and coworkers have shown that the same holds true for slowly varying delays [23]. In this model, which leaves out backtracking that would require hydrolysis of the 3' end of the transcript, addition of nucleotides is strictly sequential. Accordingly,

$$\tau_e = \sum_{i=1}^{N-1} \tau_i + k_{ae}^{-1}. \quad (32)$$

[Note that reaction (4) includes N steps, but reaction (5) only includes $N - 1$ steps, where N is the length of the transcribed

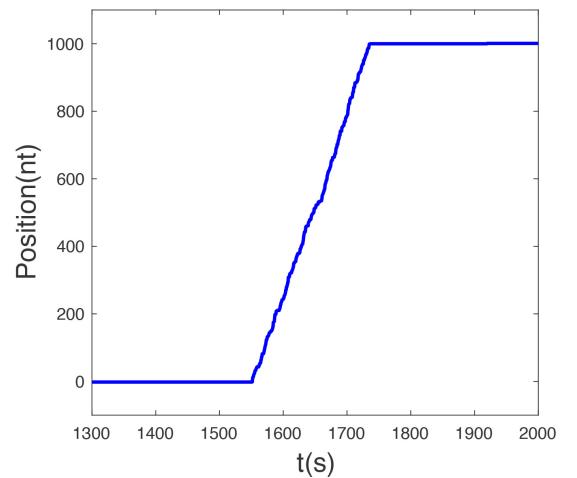


FIG. 5: Position of the RNA polymerase vs time obtained from the stochastic simulation of a single polymerase with ubiquitous pausing. The parameter values are $k_{ae} = 144 \text{ s}^{-1}$, $k_{add} = 144 \text{ s}^{-1}$, $k_{PIC} = 0.0029 \text{ s}^{-1}$, $k_{bind} = 0.0016 \text{ s}^{-1}$ and $k_{init} = 0.6 \text{ s}^{-1}$, $k_{r,i} = 3.6 \text{ s}^{-1}$, $N = 1000$ and $k_{p,i} = 100 \text{ s}^{-1}$.

sequence.] Similarly, the transition from the pre- to the post-translocated state is irreversible, so

$$\tau_i = \tau_{i,OVP} + \tau_{i,A}, \quad (33)$$

where $\tau_{i,OVP}$ is the mean residence time in the aggregated pre-translocation O and P states. As noted previously, $\tau_{i,A} = k_{add}^{-1}$. We now turn to the calculation of $\tau_{i,OVP}$.

We focus on the steps



which define the escape process from the O \vee P aggregated state. The master equation for this escape process is

$$\frac{dP_{i,P}}{dt} = k_{p,i} P_{i,O} - k_{r,i} P_{i,P}, \quad (35)$$

$$\frac{dP_{i,O}}{dt} = -k_{p,i} P_{i,O} + k_{r,i} P_{i,P} - k_{ae} P_{i,O}, \quad (36)$$

with initial conditions $P_{i,O}(0) = 1$, $P_{i,P}(0) = 0$. The probability density for the residence time in the aggregated O and P states at nucleotide i , i.e. for making a transition from state O to state A, is $\rho_{OA,i} = k_{ae} P_{i,O}(t)$. After some tedious algebra, we find

$$\rho_{OA,i} = \frac{k_{ae}}{2\sqrt{D_i}} \left[e^{-(S_i + \sqrt{D_i})t/2} \left(\sqrt{D_i} + k_{ae} + k_{p,i} - k_{r,i} \right) + e^{-(S_i - \sqrt{D_i})t/2} \left(\sqrt{D_i} - k_{ae} - k_{p,i} + k_{r,i} \right) \right], \quad (37)$$

where

$$S_i = k_{ae} + k_{p,i} + k_{r,i}, \quad (38)$$

$$D_i = S_i^2 - 4k_{ae}k_{r,i}. \quad (39)$$

As usual, we compute the mean time spent in the O and P states by

$$\tau_{i,OVP} = \int_0^\infty t \rho_{OA,i} dt = \frac{k_{p,i} + k_{r,i}}{k_{ae} k_{r,i}}. \quad (40)$$

Equation (33) now implies

$$\tau_i = \frac{k_{ae}k_{r,i} + k_{add}(k_{p,i} + k_{r,i})}{k_{ae}k_{r,i}k_{add}}. \quad (41)$$

Note that Eqs. (37) and (41) were derived without any approximations. Thus, they apply to any nucleotide, including those where long pauses occur, discussed in section IV.

It is instructive to consider a few special limits of Eq. (41). First, if there is no pausing, $k_{p,i} = 0$, and we get

$$\tau_i = \frac{k_{ae} + k_{add}}{k_{ae}k_{add}} = \frac{1}{k_{ae}} + \frac{1}{k_{add}} \quad (42)$$

which is, as expected, the sum of the mean residence times in the O and A states. On the other hand, for very strong pause sites, $k_{p,i} \gg k_{r,i}, k_{ae}$, we get

$$\tau_i \approx \frac{k_{p,i}}{k_{r,i}} \frac{1}{k_{ae}}. \quad (43)$$

In this case, the residence time in the O and P states dominates, and is computed from the pause-free O-state residence time stretched by the proportion of the time spent in the paused state.

τ_i is the mean time required to get from nucleotide i to nucleotide $i+1$. τ_i^{-1} is therefore the mean velocity of the polymerase. For the parameters of Fig. 4, we can calculate a velocity $v = 72 \text{ nt s}^{-1}$. A least-squares fit of position vs time during elongation for this one realization gives $v = 70 \text{ nt s}^{-1}$, in reasonable agreement with the theoretical value. For the case with pausing (Fig. 5), Eq. (41) leads to a velocity of

$v = 4.8 \text{ nt s}^{-1}$ while the single simulation shown in the figure gives an elongation velocity of $v = 5.4 \text{ nt s}^{-1}$. The agreement is still reasonable given the visibly larger fluctuations in the velocity in Fig. 5 than in Fig. 4, and the fact that we are only simulating a single transcription process on a gene of length 1000, which is short for a human gene [5].

C. Termination

The flux out of elongation and into termination is given by Eq. (31). Thus, the following ODEs govern the evolution during termination:

$$\frac{dP_{N,A}(t)}{dt} = J_e - k_{TC}P_{N,A}(t), \quad (44)$$

$$\frac{dP_{TC}(t)}{dt} = k_{TC}P_{N,A}(t) - k_{term}P_{TC}(t), \quad (45)$$

$$\frac{dP_T(t)}{dt} = k_{term}P_{TC}(t), \quad (46)$$

where $P_{TC}(t)$ is the probability that there is a termination complex (TC) at time t , and $P_T(t)$ represents the probability that the transcript (T) has been released, with J_e given by Eq. (31).

We can solve the ODEs based on the initial conditions of $P_{N,A}(0) = P_{TC}(0) = P_T(0) = 0$ to derive an expression for probability density of the whole transcription process, $\rho_{tr}(t)(= dP_T/dt = k_{term}P_{TC}(t))$:

$$\begin{aligned} \rho_{tr}(t) = & H(t - \tau_e) \times k_{term} k_{bind} k_{TC} k_{init} k_{PIC} \\ & \times \left([(k_{bind} - k_{init})(k_{PIC} - k_{init})(k_{term} - k_{init})(k_{TC} - k_{init})]^{-1} e^{k_{init}(t - \tau_e)} \right. \\ & + [(k_{TC} - k_{term})(k_{bind} - k_{term})(k_{init} - k_{term})(k_{PIC} - k_{term})]^{-1} e^{k_{term}(t - \tau_e)} \\ & + [(k_{bind} - k_{init})(k_{PIC} - k_{bind})(k_{bind} - k_{term})(k_{TC} - k_{bind})]^{-1} e^{k_{bind}(t - \tau_e)} \\ & + [(k_{PIC} - k_{init})(k_{PIC} - k_{bind})(k_{PIC} - k_{term})(k_{PIC} - k_{TC})]^{-1} e^{k_{PIC}(t - \tau_e)} \\ & \left. + [(k_{TC} - k_{term})(k_{TC} - k_{init})(k_{TC} - k_{bind})(k_{TC} - k_{PIC})]^{-1} e^{k_{TC}(t - \tau_e)} \right). \end{aligned} \quad (47)$$

Figures 6 and 7 show comparisons between the analytical solution and stochastic simulations of $\rho_{tr}(t)$, showing excellent agreement. Note the effect of ubiquitous pausing in Fig. 7: the shape of the distribution is not changed, but the distribution is shifted to the right with increasing pausing probability.

An alternative approach to termination, which will prove useful later, is simply to solve for the distribution of termination times. We do this by solving the equations

$$\frac{dP_{N,A}(t)}{dt} = -k_{TC}P_{N,A}(t), \quad (48)$$

$$\frac{dP_{TC}(t)}{dt} = k_{TC}P_{N,A}(t) - k_{term}P_{TC}(t), \quad (49)$$

subject to the initial conditions $P_{N,A}(0) = 1$, $P_{TC}(0) = 0$. Some elementary mathematics shows the distribution of ter-

mination times to be

$$\rho_T(t) = k_{term}P_{TC} = \frac{k_{term}k_{TC}}{k_{TC} - k_{term}} \left(e^{-k_{term}t} - e^{-k_{TC}t} \right). \quad (50)$$

IV. EFFECT OF A LONG PAUSE ALONG A GENE

The case of a site at which a lengthy pause may occur has to be treated separately, for the reasons outlined in section III B 1. We take advantage here of the modular nature of our model. Specifically, we can break up the elongation region into segments where only short ubiquitous pauses occur, dealt with using the theory of section III B and contributing only fixed delays, and segments where long pause sites are

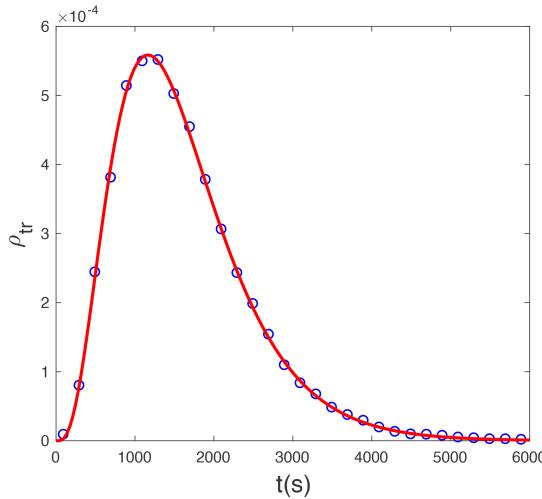


FIG. 6: A comparison between stochastic simulation (circles) and analytical (red line) solutions for $\rho_{\text{tr}}(t)$ in the absence of pausing. The parameter values are $k_{\text{ae}} = 144 \text{ s}^{-1}$, $k_{\text{add}} = 144 \text{ s}^{-1}$, $k_{\text{PIC}} = 0.0029 \text{ s}^{-1}$, $k_{\text{bind}} = 0.0016 \text{ s}^{-1}$, $k_{\text{init}} = 0.6 \text{ s}^{-1}$, $k_{\text{term}} = 0.0032 \text{ s}^{-1}$ and $N = 1000$. The numerical distribution was computed from 40 000 realizations of the transcription process.

found, as illustrated in Fig. 8. At any given long pause site, the distribution of the residence time in the combined O and P states is given by Eq. (37). We can take a convolution of this residence time distribution with the other relevant distributions to obtain the non-trivial part of the overall delay distribution, and then compute the appropriate offset due to the rest of the elongation process.

The convolution of probability densities for initiation and termination times ($\rho_{I,T}$) is calculated by

$$\rho_{I,T}(\tau) = \int_0^\tau \rho_{\text{PIC}}(\tau - t) \rho_T(t) dt. \quad (51)$$

This equation gives us an alternative path to (47): Shifting the distribution $\rho_{I,T}(t)$ by the elongation delay, τ_e gives us Eq. (47).

For a long pause at a single nucleotide at $i = i_p$, the case on which we focus here, we can obtain the convolution of $\rho_{I,T}(t)$ and $\rho_{r,P}(t)$:

$$\rho_{I,T,P}(\tau) = \int_0^\tau \rho_{OA,i_p}(\tau - t) \rho_{I,T}(t) dt. \quad (52)$$

In the event that long pauses can occur at multiple nucleotides, this last step is iterated, i.e. we take the q -fold convolution with ρ_{OA,i_p} , where q is the number of nucleotides at which long pauses may occur. Once the convolution has been calculated, the final result can be obtained by shifting the distribution by τ'_e , the delay due to all processes in elongation not involving the O \vee P aggregated state at the long pause site. The distribution for the transcription time is therefore

$$\rho_{\text{tr}} = H(t - \tau'_e) \rho_{I,T,P}(t - \tau'_e). \quad (53)$$

The expression that results from this calculation is awkward and is not reproduced here, but is available in a Maple worksheet provided in the Supplemental Materials.

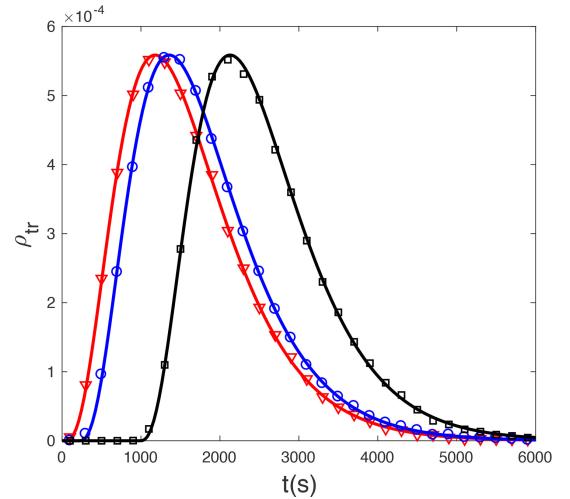


FIG. 7: Probability density for transcription time with ubiquitous pausing. We compare stochastic simulation results with analytical results. The parameter values are $k_{\text{ae}} = 144 \text{ s}^{-1}$, $k_{\text{add}} = 144 \text{ s}^{-1}$, $k_{\text{PIC}} = 0.0029 \text{ s}^{-1}$, $k_{\text{bind}} = 0.0016 \text{ s}^{-1}$ and $k_{\text{init}} = 0.6 \text{ s}^{-1}$, $k_{r,i} = 3.6 \text{ s}^{-1}$ and $N = 1000$. The pausing rate constant, $k_{p,i}$, is 7.6 s^{-1} (red curve with triangles), 100 s^{-1} (blue curve with circles) and 500 s^{-1} (black curve with squares). The numerical distribution was computed from 40 000 realizations of the transcription process.

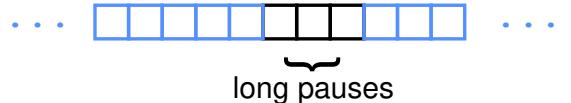


FIG. 8: Schematic illustration of a long pause region, here consisting of two nucleotides. Outside the region where long pauses may occur, elongation is treated as contributing a delay. In the region of the long pause, the residence time distribution is explicitly computed.

Let \mathcal{P} be the set of nucleotides at which a long pause may occur, and \mathcal{R} be the set of nucleotides where “regular” elongation occurs. $\mathcal{R} \cup \mathcal{S}$ is the entire transcribed sequence, i.e. \mathcal{R} is the complement of \mathcal{S} in the transcribed sequence. τ'_e includes $N - 1$ extension steps [reaction (5)], and $N - q$ steps with short pauses for which (40) applies. Thus, τ'_e is, in general,

$$\tau'_e = \frac{N - 1}{k_{\text{add}}} + \sum_{i \in \mathcal{P}} \frac{k_{p,i} + k_{r,i}}{k_{\text{ae}} k_{r,i}}. \quad (54)$$

For the specific case where long pauses only occur at one nucleotide, this reduces to $\tau'_e = (N - 1)\tau_i$, with τ_i given by Eq. (41).

Figure 9 shows the probability density of the transcription time for three different values of pausing rate constant in a case with a very long pause (mean pause time of 500 s^{-1}). Pauses of this duration might be associated with splicing, which is quite a slow process [43].

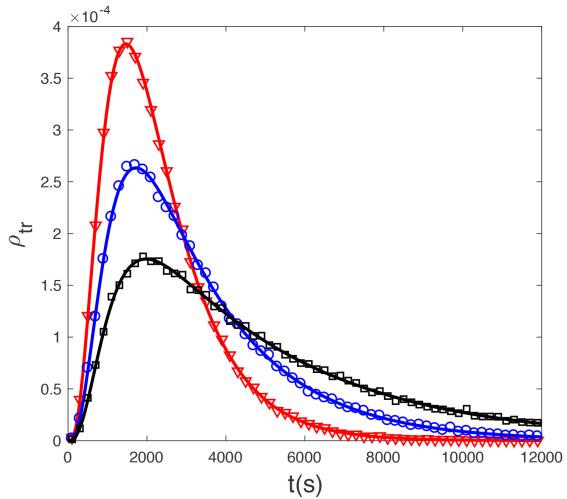
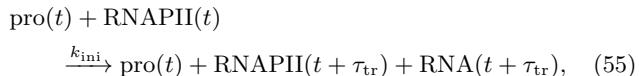


FIG. 9: Effect of a strong pause in the gene on the probability density of transcription comparing stochastic simulation results (symbols) and the analytical solution (lines). The parameter values are $k_{ae} = 144 \text{ s}^{-1}$, $k_{add} = 144 \text{ s}^{-1}$, $k_{PIC} = 0.0029 \text{ s}^{-1}$, $k_{bind} = 0.0016 \text{ s}^{-1}$ and $k_{init} = 0.6 \text{ s}^{-1}$, $k_{release} = 0.002 \text{ s}^{-1}$ and $N = 1000$. k_{pause} is 1000 s^{-1} for the black line and squares, 500 s^{-1} for the blue line and circles, and 200 s^{-1} for the red line and triangles. Each numerical distribution was computed from 10^5 realizations of the transcription process.

V. DISCUSSION AND CONCLUSIONS

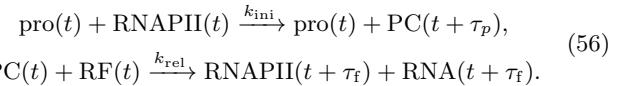
In this work, a simplified version of the transcription model proposed by Vashishtha [13] for eukaryotic cells was utilized to describe the kinetics of a single RNAPII during transcription. The master equations for initiation and termination were solved exactly. As in [23], the motion during elongation led to an advection equation so that the elongation phase has the effect of a simple delay connecting initiation and termination. Combining these results, we obtained the distribution of transcription times for genes without sites where long pauses occur. The latter situation was treated in section IV using the modularity of models of this class [20].

The delay distributions presented here are likely to be particularly useful for delay-stochastic simulations [48–52]. The delay-stochastic simulation algorithm (DSSA) is an extension of the Gillespie algorithm [53] that allows for delayed product formation. Thus, transcription can for example be written as a single delayed reaction. In the simplest case, we might write, using the delayed mass-action notation [54],



where τ_{tr} is the transcriptional delay, and we have assumed in writing this reaction that the promoter clearance time is negligible. The delay may be fixed, or we may consider the distribution of this quantity, generating a new value of the delay from the appropriate distribution each time this reaction “fires” [49]. Using the DSSA therefore requires that we know the distribution of transcription times, ρ_{tr} , if only to determine whether or not a fixed delay is appropriate.

While we have focused on the total transcription time, including the events of initiation, there are a few other ways our results could be used. We could explicitly model the initiation steps, and use a (distributed) delay to represent only elongation and termination. This would be the simplest approach to modeling the expression of a gene whose transcription is regulated at initiation. For a gene regulated at a pause site [55], elongation could be broken up into pieces, very much as we did to obtain the distributions shown in Fig. 9, treating regulation at the pause site explicitly. For example, if a release factor RF (such as P-TEFb [56]) is required to release a paused complex (PC) from an obligatory pause, we could write



If there are no other sites in the gene where long pauses occur and if the sequences preceding and following the pause site are sufficiently long, τ_p and τ_f could be treated as fixed delays. The interesting dynamics would then lie in the production and/or destruction of RF, which would itself be described by appropriate reactions, including possibly delayed reactions representing its transcription and translation.

The construction of the delay-stochastic models described above is facilitated by the modularity of the transcription model studied here. As noted by Greive and coworkers for a related model [20], this means that we can mix and match different descriptions of initiation, elongation, pausing and termination to suit any given gene. The mathematical analysis of these models is similarly modular: we can take the pieces derived here and combine them in various ways, either by convolution of distributions, or simply by a shift for the productive elongation delays.

As explained in section II, we have substantially simplified Vashishtha’s model. Studying the effects of various of the omitted details on transcription dynamics is of course a potential avenue for extending this work. The experimental observation of, as the authors put it, “deterministic” kinetics during the elongation phase of transcription [1] suggests that a more refined model of productive elongation would not behave differently than the simple model studied here. This is a consequence of the central limit theorem [57]: provided that the distribution of each τ_i has finite variance, the sum in (32) converges to a Gaussian whose standard deviation grows as \sqrt{N} while τ_e itself grows as N . In relative terms, the standard deviation of the elongation time thus decreases with N . Pausing is a different matter, as Voliotis and coworkers have shown [29]: backtracked pausing can contribute a long tail to the distribution of transcription times. To model genes that are prone to backtracking pauses, it would be possible to combine the Voliotis *et al.* pausing module with our model to obtain the corresponding distribution of elongation times.

Shortly after initiation, RNAPII tends to pause 20–40 nt downstream of the start site [58], in an event known as promoter-proximal pausing. A promoter-proximal pause that is not subject to specific regulation is relatively easy to model since it can be dealt with using the splitting trick illustrated in Fig. 8. The modularity of the model in this case means that the exact location of the pause site is irrelevant. The situation is different if the polymerase is released from its promoter-proximal pause by a specific release factor with its own dynamics. Consider again model (56). In order to use this model, we need to know the distribution of τ_p , since this

interval of time necessarily precedes the action of the release factor. However, we probably cannot treat the initial elongation leading to the pause using the theory of section III B because only a short run of nucleotides precedes promoter-proximal pausing, thus violating the central assumption in the development of the advection equation, namely that N is large. We could solve the first-passage time for reaching the pause site, but even for the moderate number of nucleotides separating the start and promoter-proximal pause sites, this is cumbersome. An additional difficulty is that the promoter-proximal pause probably occurs in a region covering several nucleotides, and not at a specific nucleotide in any given gene. Fortunately, compared to typical initiation times (Fig. 3), the contribution of the elongation time to τ_p may well be negligible despite the slower elongation rate in this region of a gene [46].

Termination is another area ripe for investigation in terms of its effect on transcription dynamics as it is quite slow in eukaryotes [4]. Indeed, we have initiated such studies in our laboratory [59]. It is also interesting in that multiple independent mechanisms cooperate to ensure termination [7], introducing an additional element of stochasticity right at the end of the process of transcription.

Even within the context of the model studied here, we have not exhausted the possibilities for further study. In particular, while we have developed the relevant equation of motion, we

have not studied in much detail the effects of variable kinetic parameters on the density of polymerases. At higher initiation rates, this could have an effect on polymerase-polymerase interactions. Some preliminary calculations on a system in which the polymerase slows down dramatically in one region of the gene have been carried out [47], but this hardly scratches the surface of what could be done.

As noted in the introduction, the eventual goal of this research is to provide delay distributions that can be used in models of gene expression. In addition to the distribution of transcription times which is the topic of this contribution, we will also need distributions for the nuclear export time, on which some progress has been made [60–62] but which remains very much an open problem, and for translation [21–23]. We can look forward to soon having all the pieces necessary either for the direct inclusion in a model of these three major steps in eukaryotic gene expression, or at the very least to enable rational decisions about whether they need to be explicitly included in a given model.

Acknowledgments

This work was funded by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

-
- [1] D. R. Larson, D. Zenklusen, B. Wu, J. A. Chao, and R. H. Singer, *Science* **332**, 475 (2011).
 - [2] S. Buratowski, S. Hahn, L. Guarente, and P. A. Sharp, *Cell* **56**, 549 (1989).
 - [3] A. Dvir, J. W. Conaway, and R. C. Conaway, *Curr. Opin. Genet. Dev.* **11**, 209 (2001).
 - [4] X. Darzacq, Y. Shav-Tal, V. de Turris, Y. Brody, S. M. Shenoy, R. D. Phair, and R. H. Singer, *Nat. Struct. Mol. Biol.* **14**, 796 (2007).
 - [5] J. C. Venter, M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, et al., *Science* **291**, 1304 (2001).
 - [6] C. N. Tennyson, H. J. Klamut, and R. G. Worton, *Nat. Genet.* **9**, 184 (1995).
 - [7] J.-F. Lemay and F. Bachand, *RNA Biol.* **12**, 927 (2015).
 - [8] U. Schmidt, E. Basyuk, M.-C. Robert, M. Yoshida, J.-P. Villemain, D. Auboeuf, S. Aitken, and E. Bertrand, *J. Cell Biol.* **193**, 819 (2011).
 - [9] P. Cramer, A. Srebrow, S. Kadener, S. Werbajh, M. de la Mata, G. Melen, G. Nogués, and A. R. Kornblith, *FEBS Lett.* **498**, 179 (2001).
 - [10] A. Babour, C. Dargemont, and F. Stutz, *Biochim. Biophys. Acta* **1819**, 521 (2012).
 - [11] M. R. Roussel and R. Zhu, *Bull. Math. Biol.* **68**, 1681 (2006).
 - [12] M. R. Roussel, *BIOMATH* **2**, 1307247 (2013).
 - [13] S. Vashishtha, M.Sc. thesis, University of Lethbridge (2011), URL <https://www.uleth.ca/dspace/handle/10133/3190>.
 - [14] V. Pelechano, S. Chávez, and J. E. Pérez-Ortíz, *PLoS One* **5**, e15442 (2010).
 - [15] T. D. Yager and P. H. Von Hippel, *Biochemistry* **30**, 1097 (1991).
 - [16] F. Jülicher and R. Bruinsma, *Biophys. J.* **74**, 1169 (1998).
 - [17] H.-Y. Wang, T. Elston, A. Mogilner, and G. Oster, *Biophys. J.* **74**, 1186 (1998).
 - [18] L. Bai, A. Shundrovsky, and M. D. Wang, *J. Mol. Biol.* **344**, 335 (2004).
 - [19] L. Bai, R. M. Fulbright, and M. D. Wang, *Phys. Rev. Lett.* **98**, 068103 (2007).
 - [20] S. J. Greive, J. P. Goodarzi, S. E. Weitzel, and P. H. von Hippel, *Biophys. J.* **101**, 1155 (2011).
 - [21] A. Garai, D. Chowdhury, D. Chowdhury, and T. V. Ramakrishnan, *Phys. Rev. E* **80**, 011908 (2009).
 - [22] A. Garai, D. Chowdhury, and T. V. Ramakrishnan, *Phys. Rev. E* **79**, 011916 (2009).
 - [23] L. Mier-y-Terán-Romero, M. Silber, and V. Hatzimanikatis, *PLoS Comput. Biol.* **6**, e1000726 (2010).
 - [24] K. C. Neuman, E. A. Abbondanzieri, R. Landick, J. Gelles, and S. M. Block, *Cell* **115**, 437 (2003).
 - [25] L. S. Churchman and J. S. Weissman, *Nature* **469**, 368 (2011).
 - [26] R. Landick, *Biochem. Soc. Trans.* **34**, 1062 (2006).
 - [27] V. Epshtein, F. Toulmé, A. R. Rahmouni, S. Borukhov, and E. Nudler, *EMBO J.* **22**, 4719 (2003).
 - [28] S. Klumpp, *J. Stat. Phys.* **142**, 1252 (2011).
 - [29] M. Voliotis, N. Cohen, C. Molina-París, and T. B. Liverpool, *Biophys. J.* **94**, 334 (2008).
 - [30] P. B. Rahl, C. Y. Lin, A. C. Seila, R. A. Flynn, S. McCuine, C. B. Burge, P. A. Sharp, and R. A. Young, *Cell* **141**, 432 (2010).
 - [31] J. Li and D. S. Gilmour, *Curr. Opin. Genet. Dev.* **21**, 231 (2011).
 - [32] S. Nechaev and K. Adelman, *Biochim. Biophys. Acta* **1809**, 34 (2011).
 - [33] B. Zamft, L. Bintu, T. Ishibashi, and C. Bustamante,

- Proc. Natl. Acad. Sci. U.S.A. **109**, 8948 (2012).
- [34] R. D. Alexander, S. A. Innocente, J. D. Barrass, and J. D. Beggs, Mol. Cell **40**, 582 (2010).
- [35] N. Gromak, S. West, and N. J. Proudfoot, Mol. Cell. Biol. **26**, 3986 (2006).
- [36] D. S. Luse, Biochim. Biophys. Acta **1829**, 63 (2013).
- [37] H.-J. Woo, Phys. Rev. E **74**, 011907 (2006).
- [38] M. H. Larson, J. Zhou, C. D. Kaplan, M. Palangat, R. D. Kornberg, R. Landick, and S. M. Block, Proc. Natl. Acad. Sci. U.S.A. **109**, 6555 (2012).
- [39] A. C. M. Cheung and P. Cramer, Nature **471**, 249 (2011).
- [40] G. Brzyżek and S. Świeżewski, Transcription **6**, 37 (2015).
- [41] M. Imashimizu, M. L. Kireeva, L. Lubkowska, D. Gotte, A. R. Parks, J. N. Stratheim, and M. Kashlev, J. Mol. Biol. **425**, 697 (2013).
- [42] J. W. Roberts, Science **344**, 1226 (2014).
- [43] J. Singh and R. A. Padgett, Nat. Struct. Mol. Biol. **16**, 1128 (2009).
- [44] E. Rosonina, S. Kaneko, and J. L. Manley, Genes Dev. **20**, 1050 (2006).
- [45] R. A. Coleman and B. F. Pugh, Proc. Natl. Acad. Sci. U.S.A. **94**, 7221 (1997).
- [46] H. A. Ferguson, J. F. Kugel, and J. A. Goodrich, J. Mol. Biol. **314**, 993 (2001).
- [47] S. H. Hosseini, Master's thesis, University of Lethbridge (2016), URL <https://www.uleth.ca/dspace/handle/10133/4791>.
- [48] D. Bratsun, D. Volfson, L. S. Tsimring, and J. Hasty, Proc. Natl. Acad. Sci. U.S.A. **102**, 14593 (2005).
- [49] M. R. Roussel and R. Zhu, Phys. Biol. **3**, 274 (2006).
- [50] A. S. Ribeiro and J. Lloyd-Price, Bioinformatics **23**, 777 (2007).
- [51] J. Lloyd-Price, A. Gupta, and A. S. Ribeiro, Bioinformatics **28**, 3004 (2012).
- [52] T. Maarleveld, *StochPy User Guide, Release 2.3.0* (2015), URL https://sourceforge.net/projects/stochpy/files/stochpy_userguide_2.3.pdf/ download.
- [53] D. T. Gillespie, J. Comput. Phys. **22**, 403 (1976).
- [54] M. R. Roussel, J. Phys. Chem. **100**, 8323 (1996).
- [55] R. J. Sims III, R. Belotserkovskaya, and D. Reinberg, Genes Dev. **18**, 2437 (2004).
- [56] B. H. Jennings, BioEssays **35**, 553 (2013).
- [57] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam, 1981).
- [58] B. Li, J. A. Weber, Y. Chen, A. L. Greenleaf, and D. S. Gilmour, Mol. Cell. Biol. **16**, 5433 (1996).
- [59] R.-J. Murphy, Master's thesis, University of Lethbridge (2017), URL <https://www.uleth.ca/dspace/handle/10133/4906>.
- [60] S. Tang, Ph.D. thesis, University of Lethbridge (2010), URL <https://www.uleth.ca/dspace/handle/10133/2567>.
- [61] M. R. Roussel and T. Tang, IET Syst. Biol. **6**, 125 (2012).
- [62] D. Holcman and Z. Schuss, SIAM Rev. **56**, 213 (2014).