THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

Cite This: J. Phys. Chem. B 2018, 122, 11969−11977

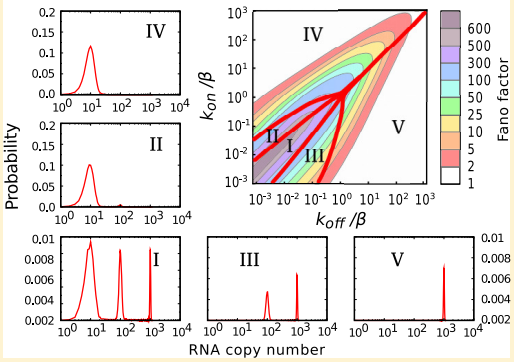
pubs.acs.org/JPCB

Theoretical Investigation of Transcriptional Bursting: A Multistate Approach

Alena Klindziuk[†,‡] and Anatoly B. Kolomeisky*[,†,‡,§]

[†]Department of Chemistry, [‡]Center for Theoretical Biological Physics, and [§]Department of Chemical and Biomolecular Engineering, Rice University, Houston, Texas 77005, United States

ABSTRACT: Variability in gene expression causes genetically identical cells to exhibit different phenotypes. One probable cause of this variability is transcriptional bursting, where the synthesis of RNA molecules randomly alternates with periods of silence in the transfer of genetic information. Yet, the molecular mechanisms behind this variability remain unclear. Experiments indicate that multiple biochemical states might be involved in the production of RNA molecules. Stimulated by these observations, we developed a theoretical framework to investigate the mechanisms of transcriptional bursting. It is based on a multistate stochastic approach that provides a full quantitative description of the dynamic properties in the system. We found that the degree of stochastic fluctuations during transcription directly correlates with the number of biochemical states. This explains experimentally observed variability and fluctuations in the quantities of the produced RNA molecules. The procedure to estimate the number of relevant biochemical states participating in the transcription is outlined and applied for analysis of experimental results. We also developed a general dynamic phase diagram for the transcription process. The presented theoretical method clarifies physical−chemical aspects of the transcriptional bursting and presents a minimal chemical-kinetic description of the process.



INTRODUCTION

Transcription is one of the most fundamental processes in nature. It consists of copying the genetic information contained in DNA into complementary RNA molecules, which eventually leads to proper development, function, and regulation of cellular organisms.[1] Transcription involves a complex network of biochemical and biophysical processes, which maintains its robustness in all living systems at all times. Although significant progress in explaining how the transcription works has been achieved,[1] many aspects of its molecular mechanisms still remain poorly understood.

It is known that cells with identical genomes never exhibit exactly the same physical characteristics.[1] This variability in gene expression is observed in all cell systems ranging from the bacterial colonies to specialized tissue cells of complex multicellular organisms.[1−3] However, the molecular origin of this gene expression noise is not completely understood yet.[3] It has been suggested that transcription, which is the first step of gene expression, is the dominating factor behind the molecular fluctuations in gene expression.[3,4] These arguments are supported by experiments exhibiting transcriptional bursting phenomena when the continuous production of RNA molecules is interrupted by periods of inactivity.[5,6] This stochasticity of transcription dynamics has been observed in prokaryotes, yeast, and mammalian cells,[4,7,8] and it is commonly described by a minimal two-state model.[5,6,9,10] In this view, the transcription of a gene randomly switches between ON and OFF states. In the ON state, the RNA molecules are generated and degraded via a birth and death process, whereas in the OFF state, no transcripts are produced and the degradation happens as in the ON state.

Transcriptional bursting has been intensively investigated in recent years using both experimental and theoretical methods.[4−25] Many features of this phenomenon are now better understood.[5] However, there are still questions about the microscopic origin of the transcriptional bursting and its influence on the dynamic properties of the biochemical system. The distributions of produced RNA molecules exhibit large variability in different systems.[6,10,14] In some cases, it gives a purely monotonic behavior (peaked at zero), whereas in other cases, one or several peaks appear in the distribution.[6,12] Yet, the molecular mechanism of this variability remains unclear. In addition, experiments show a large variability in stochastic fluctuations during RNA production, which cannot be easily connected with the number of produced RNA molecules.[12] Furthermore, recent experiments indicate that more than two biochemical states might participate in the transcription process, although it is not clear how to determine the number of relevant states from experimental data.[7,8,17−19] At the same time, current theoretical views on the mechanisms of transcription bursting mostly employ two-state kinetic models.[12] There are several studies of transcription bursting that employ multistate models.[24,25] In ref 24, the stationary-state distributions of biochemical states during the tran-

Received: October 3, 2018
Revised: November 23, 2018
Published: November 27, 2018

DOI: 10.1021/acs.jpcb.8b09676
J. Phys. Chem. B 2018, 122, 11969−11977

scription are directly calculated using a spectral method. However, this mathematically quite elaborate approach provides only numerical solutions and it does not calculate the average properties reported in experiments. In ref 25, the multistate model of transcription bursting is analyzed analytically. However, this approach assumes that there is only one ON state and multiple OFF states in the system, which do not agree with recent observations of multiple production states during transcription.[7,8,17−19] This suggests that more advanced theoretical models are needed to describe a wider range of transcription regulation scenarios.[7]

Stimulated by these observations, in this paper, we develop a multistate chemical-kinetic approach to analyze the molecular mechanisms of transcriptional bursting. Our goal is to build a minimal theoretical description that provides a clear molecular picture of the underlying processes and which is also consistent with experimental observations. The discrete-state stochastic method allows us to evaluate dynamic properties of the transcription process with multiple switches between various biochemical states. Several simple models of the transcription process are explicitly analyzed and discussed. A general dynamic phase diagram of different behaviors in transcription is presented and explained. It is shown that the number of peaks in the distributions of produced RNA molecules is specified by the number of independent biochemical states, which is determined by the dominating chemical transitions in the system. Here, independent biochemical states are defined as those in which the system spends significant periods of time during the transcription and which are specified by distinct sets of chemical-kinetic transition rates. The number of independent biochemical states might be equal or less than the number of actual kinetic states in the system. We also find that the degree of stochastic variations in the number of RNA molecules directly correlates with the possibility of exploring various biochemical states; i.e., larger fluctuations correspond to the systems with more independent biochemical states. In addition, our method provides a simple approach to evaluate the number of relevant biochemical states in the system. Analytical calculations are fully supported by computer Monte-Carlo simulations.

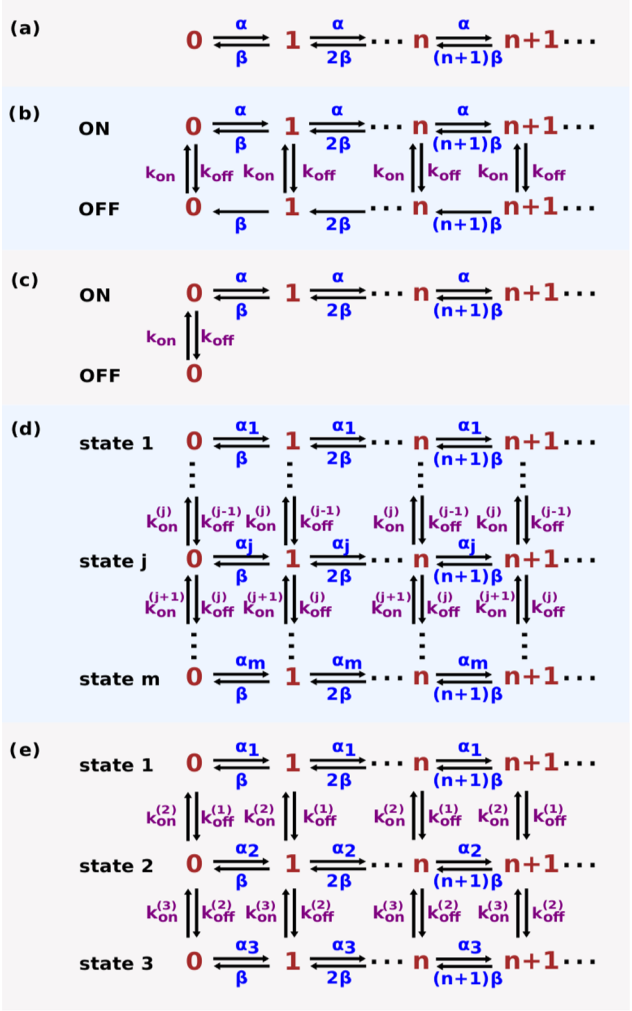


Figure 1. Chemical-kinetic schemes for multistate models: (a) one-state model, (b) two-state model, (c) "Poisson with zero spike" model, (d) general multistate model, and (e) three-state model. Details of the models are explained in the text.

THEORETICAL MODELS AND RESULTS

One-State Model. To understand the origin of variability and stochastic fluctuations in transcription, it is important to consider the simplest one-state model presented in Figure 1a, whereas different multistate models are illustrated in Figure 1b−e. In this model, there is only one biochemical state (represented as a single chain of connected microstates) in which the RNA molecules are continuously synthesized with a rate α and degraded with a first-order rate constant β (see Figure 1a) so that the production rate remains constant, while the degradation rate is proportional to the number of existing RNA transcripts in the system. Since the number of RNA transcripts n can be very low, a stochastic description of the dynamics of the system is required.[12] This model has been extensively investigated before, and the reason we present it here is to better explain how the existence of multiple kinetic biochemical states affects the statistics of RNA production during transcription.

We define $P_n(t)$ as the probability of the system to have n RNA transcripts at time t. Let us assume that the system reaches the stationary state and $P_n(t \to \infty) \equiv P_n$. The probability flux to move from a microstate n to a microstate $n + 1$ is given by αP_n, and the reverse flux from $n + 1$ to n is equal to $(n + 1)\,\beta P_{n+1}$. Because the system can be viewed as a single chain of sequential microstates (see Figure 1a), the overall flux between every two neighboring microstates in the stationary state must be zero, which leads to

$$\alpha P_n = \beta(n + 1)P_{n+1} \tag{1}$$

Combining this result with the normalization condition $(\sum_{n=0}^{\infty} P_n = 1)$ gives explicit expression for the stationary distribution of produced RNA molecules in the one-state model

$$P_n = \frac{x^n\, e^{-x}}{n!} \tag{2}$$

where x is an equilibrium constant for the synthesis/degradation process defined as $x = \alpha/\beta$. This is a well-known Poisson distribution, which is illustrated in Figure 2a.[6]

The explicit expression for the distribution allows us to evaluate all relevant dynamic properties of the RNA production in this model. First, the average number of RNA transcripts, $\langle n \rangle$, can be calculated as

DOI: 10.1021/acs.jpcb.8b09676
J. Phys. Chem. B 2018, 122, 11969−11977

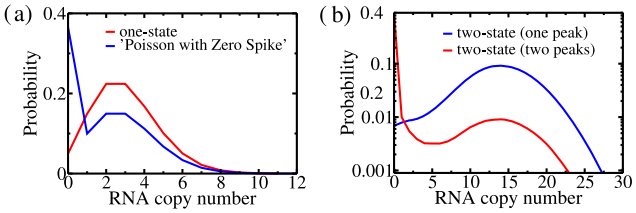


Figure 2. Stationary distributions of RNA molecules in various models: (a) one-state and Poisson with zero spike models. Parameters used for calculations are: $\alpha = 30$, $\beta = 10$, $k_{on} = 10$, and $k_{off} = 100$. (b) Two-state models. Parameters used for calculations are: $\alpha = 15$, $\beta = 1$, $k_{on} = 0.01$, and $k_{off} = 0.1$ for distribution with two peaks and $\alpha = 15$, $\beta = 1$, $k_{on} = 1$ and $k_{off} = 0.1$ for distribution with one peak.

$$\langle n \rangle = \sum_{n=0}^{\infty} nP_n = x \tag{3}$$

Thus, the average number of RNA molecules in the system is given by the equilibrium constant for the RNA synthesis/degradation process. Similarly, one can calculate other moments of the distribution of produced RNA molecules. The second moment is given by

$$\langle n^2 \rangle = \sum_{n=0}^{\infty} n^2 P_n = x^2 + x \tag{4}$$

From eqs 3 and 4, one can estimate the distribution variance σ^2

$$\sigma^2 = \langle n^2 \rangle - \langle n \rangle^2 = x \tag{5}$$

To quantify the degree of stochastic fluctuations, it is convenient to introduce a dimensionless parameter F, known as the Fano factor, which is defined as $F = \sigma^2/\langle n \rangle$. It is easy to show that in the one-state model, we always have $F = 1$, and this is the signature of the Poisson distribution. Because experiments clearly show deviations from the Poisson distribution,[6,10,15] one can see that multistate models are needed to properly describe the dynamics and fluctuations in the transcription process.

General Multistate Model. Let us consider a general multistate model for the transcription, as presented in Figure 1d. It is important to note here that, following our goal of presenting a minimalist theoretical framework, in this approach, we do not take into account various feedback mechanisms of genetic regulation, which might be important.[26,27] The system can be found in one of m kinetic biochemical states (m chains of microstates), and the production rate in the state j ($j = 1,..., m$) is equal to α_j, whereas the degradation rate constant is β; i.e., it is the same in all states. We define $k_{off}^{(j)}$ as the transition rate from the state j to $j + 1$ and $k_{on}^{(j)}$ as the transition rate from the state j to $j - 1$ (Figure 1d). It is convenient also to define $x_j = \alpha_j/\beta$ and $\gamma_j = k_{off}^{(j)}/k_{on}^{(j+1)}$ as equilibrium constants for RNA synthesis/degradation in the state j and for the switching transitions between the states j and $j + 1$, respectively.

We define $P_n^{(j)}(t)$ ($j = 1,..., m$) as the probability of being in the biochemical state j with n RNA molecules at time t. The temporal evolution of the system can be described by a system of m master equations for $n \geq 1$

$$\frac{dP_n^{(1)}(t)}{dt} = \alpha_1 P_{n-1}^{(1)}(t) + k_{on}^{(2)} P_n^{(2)}(t) + (n+1)\beta P_{n+1}^{(1)}(t) - (n\beta + k_{off}^{(1)} + \alpha_1) P_n^{(1)}(t) \tag{6}$$

$$\frac{dP_n^{(j)}(t)}{dt} = \alpha_j P_{n-1}^{(j)}(t) + k_{on}^{(j+1)} P_n^{(j+1)}(t) + k_{off}^{(j-1)} P_n^{(j-1)}(t) + (n+1)\beta P_{n+1}^{(j)}(t) - (n\beta + k_{off}^{(j)} + k_{on}^{(j)} + \alpha_j) P_n^{(j)}(t) \tag{7}$$

$$\frac{dP_n^{(m)}(t)}{dt} = \alpha_m P_{n-1}^{(m)}(t) + k_{off}^{(m-1)} P_n^{(m-1)}(t) + (n+1)\beta P_{n+1}^{(m)}(t) - (n\beta + k_{on}^{(m)} + \alpha_m) P_n^{(m)}(t) \tag{8}$$

For $n = 0$, the master equations have a slightly simpler form because the number of RNA transcripts cannot be less than zero

$$\frac{dP_0^{(1)}(t)}{dt} = k_{on}^{(2)} P_0^{(2)}(t) + \beta P_1^{(1)}(t) - (k_{off}^{(1)} + \alpha_1) P_0^{(1)}(t) \tag{9}$$

$$\frac{dP_0^{(j)}(t)}{dt} = k_{on}^{(j+1)} P_0^{(j+1)}(t) + k_{off}^{(j-1)} P_0^{(j-1)}(t) + \beta P_1^{(j)}(t) - (k_{off}^{(j)} + k_{on}^{(j)} + \alpha_j) P_0^{(j)}(t) \tag{10}$$

$$\frac{dP_0^{(m)}(t)}{dt} = k_{off}^{(m-1)} P_0^{(m-1)}(t) + \beta P_1^{(m)}(t) - (k_{on}^{(m)} + \alpha_m) P_0^{(m)}(t) \tag{11}$$

The stationary-state limit occurs when $\frac{dP_n^{(j)}(t)}{dt} = 0$ and $P_n^{(j)}(t) \equiv P_n^{(j)}$.

To obtain a dynamic description of the RNA production in this system, the generating function method is employed.[9] More specifically, we define a set of new functions $g_j(z)$ ($j = 1,..., m$)

$$g_j(z) = P_0^{(j)} + zP_1^{(j)} + z^2 P_2^{(j)} + \cdots = \sum_{n=0}^{\infty} z^n P_n^{(j)} \tag{12}$$

It can be shown that

$$zg_j' = \sum_{n=0}^{\infty} nz^n P_n^{(j)} \tag{13}$$

where $g_j' \equiv \frac{dg_j(z)}{dz}$ and

$$z^2 g_j'' + zg_j' = \sum_{n=0}^{\infty} n^2 z^n P_n^{(j)} \tag{14}$$

In addition, from eq 12, we obtain

$$g_j(z=1) = \sum_{n=0}^{\infty} P_n^{(j)} \tag{15}$$

and the normalization condition gives

$$\sum_{j=1}^{m} \sum_{n=0}^{\infty} P_n^{(j)} = \sum_{j=1}^{m} g_j(z=1) = 1 \tag{16}$$

Then, the master equations can be rewritten in terms of the generating functions as

$$\beta y g_1'(y) = (\alpha_1 y - k_{off}^{(1)}) g_1(y) + k_{on}^{(2)} g_2(y) \tag{17}$$

DOI: 10.1021/acs.jpcb.8b09676
J. Phys. Chem. B 2018, 122, 11969−11977

$$\beta y g_j'(y) = (\alpha_j y - k_{off}^{(j)} - k_{on}^{(j)}) g_j(y) + k_{off}^{(j-1)} g_{j-1}(y) + k_{on}^{(j+1)} g_{j+1}(y) \quad (18)$$

$$\beta y g_m'(y) = (\alpha_m y - k_{on}^{(m)}) g_m(y) + k_{off}^{(m-1)} g_{m-1}(y) \quad (19)$$

where $y = z - 1$. The generating functions can be expanded around $y = 0$, producing

$$g_j(y) \simeq d_j + f_j y + \frac{1}{2} h_j y^2 + \cdots \quad (20)$$

The unknown coefficients d_j, f_j, and h_j can be found by substituting the expansion into eqs 17−19 and balancing terms of the same order of y on both sides of these equations. This yields

$$d_j(k_{off}^{(j)} + k_{on}^{(j)}) = k_{off}^{(j-1)} d_{j-1} + k_{on}^{(j+1)} d_{j+1} \quad (21)$$

$$f_j(\beta + k_{off}^{(j)} + k_{on}^{(j)}) = \alpha_j d_j + k_{off}^{(j-1)} f_{j-1} + k_{on}^{(j+1)} f_{j+1} \quad (22)$$

$$h_j(2\beta + k_{off}^{(j)} + k_{on}^{(j)}) = 2\alpha_j f_j + k_{off}^{(j-1)} h_{j-1} + k_{on}^{(j+1)} h_{j+1} \quad (23)$$

The first and second moments of the distributions can be expressed in terms of these coefficients, leading to

$$\langle n \rangle = (g'_1 + \cdots + g'_m)|_{y=0} = f_1 + \cdots + f_m = \sum_{j=1}^{m} f_j = \sum_{j=1}^{m} x_j d_j \quad (24)$$

$$\langle n^2 \rangle = (g''_1 + g'_1 + \cdots + g''_m + g'_m)|_{y=0} = \sum_{j=1}^{m} f_j + \sum_{j=1}^{m} h_j = \sum_{j=1}^{m} x_j d_j + \sum_{j=1}^{m} x_j f_j \quad (25)$$

Thus, the stationary properties of the system can be explicitly evaluated if we know the relations between the coefficients d_j, f_j, and h_j and the transition rates. By taking into account the normalization condition when eq 16 leads to $\sum_{j=1}^{m} d_j = 1$, eq 21 can be solved to produce

$$d_j = \frac{\prod_{i=1}^{j-1} \gamma_i}{1 + \sum_{k=1}^{m-1} \prod_{i=1}^{k} \gamma_i} \quad (26)$$

where $\gamma_j = k_{off}^{(j)}/k_{on}^{(j+1)}$, as defined above. This allows us to evaluate $\langle n \rangle$ at general conditions from eq 24. We could not obtain a compact general expression for f_j but for any specific value of m (number of states in the system), the explicit calculations can be done using matrix equations, as we show below in detail for $m = 3$. At the same time, some general results can be obtained in the limiting cases. If the transition rates between states are small, $k_{on}^{(j)}$, $k_{off}^{(j)} \ll \alpha_j$, β, then from eq 22, we have $f_j \approx x_j d_j$, which leads to

$$\langle n^2 \rangle = \sum_{j=1}^{m} (x_i^2 + x_j) d_j \quad (27)$$

In this limit, the expression for the Fano factor is given by

$$F = 1 + \frac{\sum_{j=1}^{m} x_j^2 d_j - \left(\sum_{j=1}^{m} x_j d_j \right)^2}{\sum_{j=1}^{m} x_j d_j} \quad (28)$$

which corresponds to large stochastic fluctuations (larger than those for the one-state model). In the limit when the switching rates are very large, $k_{on}^{(j)}$, $k_{off}^{(j)} \gg \alpha_j$, β, we obtain

$$\langle n^2 \rangle = \left(\sum_{j=1}^{m} x_j d_j \right)^2 + \sum_{j=1}^{m} x_j d_j \quad (29)$$

and

$$F = \frac{\sigma^2}{\langle n \rangle} = 1 \quad (30)$$

In this case, the stochastic fluctuations are minimal and the system can be viewed as an effective one-state model with properly renormalized transition rates.

To understand the transcriptional bursting, we have to discuss the meaning of states in our analysis. Here, we distinguish kinetic biochemical states from independent biochemical states. Each kinetic biochemical state is specified by a unique production rate and a set of transition rates, some of which may be the same. The independent biochemical states are defined as the states that are visited for significant periods of time at stationary-state conditions and which have distinct sets of transition rates. Clearly, the number of independent biochemical states could be the same or smaller than the number of kinetic biochemical states and it depends on the amplitudes of transition rates. For example, if switching rates to a specific state are large in comparison with other transitions, the system will effectively explore only this state and the number of independent states is equal to 1.

The general results for the degree of variability in RNA production in both limiting cases can be explained using the following arguments: for slow transition rates, there are generally m independent biochemical states and this leads to large stochastic fluctuations since the system can explore all of them in the stationary-state limit. For fast transition rates, there is an effective equilibrium between all biochemical states involved in transcription, which can be viewed as having only one effective independent "state" in the system. In this case, the situation is similar to the one-state model with Poisson distribution where the stochastic fluctuations are minimal. For intermediate transition rates, various dynamic behaviors are expected depending on the number of independent biochemical states (states with distinct chemical-kinetic transition rates where the system spends significant periods of time), as we show below for the three-state model. Thus, we argue that the amplitude of stochastic fluctuations in the transcription correlates with the number of relevant biochemical states in the system.

Our analytical results for the general multistate model also provide insights into how to evaluate the number of independent biochemical states from the experimental data. To explore all states m, the system needs to have slow transition rates so that each state can be visited for significant periods of time. If we assume first that transition rates between states are comparable, i.e., $k_{on}^{j} \simeq k_{off}^{j} \ll \alpha_j$ and β, then we can estimate that $d_j \simeq 1/m$ and employ eq 28. If all equilibrium constants for synthesis/degradation in all states are also comparable, $x_j \simeq x$, then we obtain

DOI: 10.1021/acs.jpcb.8b09676
J. Phys. Chem. B 2018, 122, 11969−11977

$$\langle n\rangle \simeq x,\ F \simeq 1 \tag{31}$$

This result is easy to understand. All chemical-kinetic rates for each state are comparable, and this means that these states are not distinguishable, leading to an effectively single biochemical state with the expected Poisson dynamics.

A more realistic situation is when at least one of the equilibrium constants for synthesis/degradation is larger than the others, i.e., for a specific state j, we have $x_j = x \gg x_{i\neq j}$. Then, our calculations yield

$$\langle n\rangle \simeq x/m,\ F \simeq 1 + x(1 - 1/m) \tag{32}$$

The last equation can be rewritten as

$$m \simeq 1 + (F - 1)/\langle n\rangle \tag{33}$$

This is an important result because it shows how to obtain the information on underlying chemical-kinetic network of states. From the experimentally measured values of the average number of RNA transcripts and the Fano factor, one can estimate the number of relevant biochemical states in the system. Because our arguments are presented for the situation with largest possible fluctuations in the system, eq 33 evaluates the number of independent states that can explain the observed transcription dynamics and thus it gives the minimal estimate on the number of kinetic biochemical states. For example, measurements of transcriptional bursting in *Escherichia coli* bacteria gave $\langle n\rangle \simeq 10$ and $F \simeq 4.1$,[6] which after substitution into eq 33 suggests that $m \simeq 2$. Similar results (see Figure 3) can be obtained for experimental data on other genes

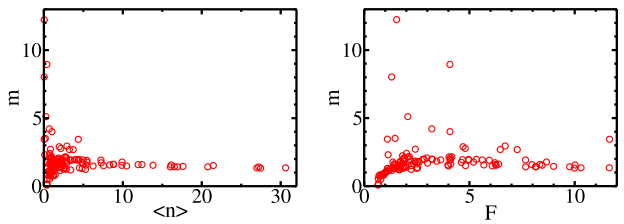


Figure 3. Estimate of the minimal number of independent biochemical states m from corresponding experimental measurements of F and $\langle n\rangle$ for different genes in *E. coli* (data from ref 10).

in this bacteria.[10] It shows that for *E. coli* bacteria, the two-state chemical-kinetic network explains well the observations for most genes. This means that the two-state description is generally reasonable for this system and there is no need to invoke more biochemical states on the basis of these experimental data. However, the situation might be different in other organisms. In addition, the presence of feedback mechanisms might also affect our conclusions on the number of relevant states in transcriptional bursting.[26,27]

Two-State Model. To illustrate how our general approach works, let us examine a specific model that has been already used to describe transcriptional bursting in the past. We start with the two-state model that features an ON state and an OFF state, as presented in Figure 1b. In the ON state, the RNA molecules are synthesized with a rate α and degraded with a rate constant β. The system can stochastically switch to the OFF state with a rate k_{off} where degradation with the rate constant β is only taking place. From the OFF state, it can also transition back to the active state with a rate k_{on} (see Figure 1b). This model has been theoretically analyzed before.[9] For this reason, we only briefly present the main results here, emphasizing that this model is a special case of our more general multistate description.

Two generating functions can be introduced for the stationary-state analysis of this model[9]

$$g_1(z) = \sum_{n=0}^{\infty} P_n^{(1)},\ g_2(z) = \sum_{n=0}^{\infty} P_n^{(2)} \tag{34}$$

We expanded them around $y = 0$ $(y = z - 1)$, producing

$$g_1(y) \approx d_1 + f_1 + \frac{1}{2}h_1 y^2 + \cdots \tag{35}$$

and

$$g_2(y) \approx d_2 + f_2 + \frac{1}{2}h_2 y^2 + \cdots \tag{36}$$

The expansion can be substituted back into master equations to determine the unknown coefficients d_1, d_2, f_1, f_2, h_1, and h_2. From this procedure, we obtain

$$d_1 = \frac{1}{1+\gamma},\ d_2 = \frac{\gamma}{1+\gamma} \tag{37}$$

and

$$f_1 = \frac{\alpha}{(1+\gamma)}\frac{(\beta + k_{on})}{(\beta + k_{on} + k_{off})},$$
$$f_2 = \frac{\alpha}{(1+\gamma)}\frac{k_{off}}{(\beta + k_{on} + k_{off})} \tag{38}$$

In addition, it can be shown that

$$(h_1 + h_2) = xf_1 \tag{39}$$

The types of distributions of RNA transcripts that are possible in the two-state model are presented in Figure 2b. Depending on the kinetic parameters, one or two peaks might be observed in the distributions of the produced RNA molecules. Now, we can determine the first and the second moments of the distribution

$$\langle n\rangle = \sum_{n=0}^{\infty} n(P_n^{(1)} + P_n^{(2)}) = (g_1' + g_2')|_{y=0} = f_1 + f_2 \tag{40}$$

$$\langle n^2\rangle = \textstyle\sum_{n=0}^{\infty} n^2(P_n^{(1)} + P_n^{(2)}) = (g''_1 + g''_2)|_{y=0}$$
$$= h_1 + h_2 + f_1 + f_2 \tag{41}$$

Using eqs 37−39, we obtain

$$\langle n\rangle = \frac{x}{1+\gamma} \tag{42}$$

$$\langle n^2\rangle = \frac{x^2}{(1+\gamma)}\frac{(\beta + k_{on})}{(\beta + k_{on} + k_{off})} + \frac{x}{1+\gamma} \tag{43}$$

Finally, for the Fano factor, our calculations yield

$$F = \frac{\langle n^2\rangle - \langle n\rangle^2}{\langle n\rangle} = 1 + \frac{x\gamma}{(1+\gamma)}\frac{\beta}{(\beta + k_{off} + k_{on})} \tag{44}$$

The results of the calculations are presented in Figure 4. Adding a second state where only degradation is taking place lowers the mean number of the produced RNA molecules and increases the stochastic fluctuations. One can see again that the largest fluctuations are achieved for slow switching rates between biochemical states. Increasing the switching frequency

DOI: 10.1021/acs.jpcb.8b09676
J. Phys. Chem. B 2018, 122, 11969−11977

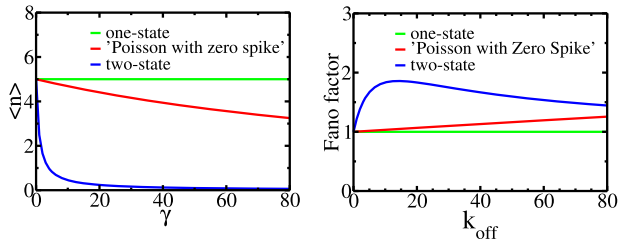


Figure 4. First moments and the Fano factors for the distributions of RNA copies in various models. Left panel: mean number of RNA molecules as a function of the state transition equilibrium constant. Parameters used for calculations are: $x = 5$. Right panel: the Fano factor as a function of the rate k_{off}. Parameters used for calculations are: $\alpha = 50$, $\beta = 10$, and $k_{on} = 10$.

lowers the fluctuations because the system can now be viewed as being in effectively one biochemical state.

Poisson with Zero Spike Model. To describe experimental data from transcription, a simpler model, called Poisson with zero spike, has been introduced.[13] It is again assumed that the system can be found in an ON or OFF state, as shown in Figure 1c. However, although any number of microstates can be observed in the ON state, the OFF state can only have zero RNA molecules. It has been argued that this approximate description is valid for slow transition rates between the states.[13]

As before, $P_n^{(j)}(t)$ is the probability of finding the system in the state j ($j = 1$ is for the ON state, and $j = 2$ is for the OFF state) at time t having n RNA molecules. In the ON state, n can take any value, whereas in the OFF state, $n = 0$ is the only possibility. Considering the stationary-state limit, it was found that[13]

$$P_0^{(1)} = \frac{e^{-x}}{(1+\gamma)},\quad P_0^{(2)} = \frac{\gamma}{(1+\gamma)},\quad P_{n\geq 1}^{(1)} = \frac{e^{-x}}{(1+\gamma)}\frac{x^n}{n!} \tag{45}$$

It was also shown that the average number of produced RNA molecules and the Fano factor are given by[13]

$$\langle n\rangle = \frac{x}{(1+\gamma)},\quad F = 1 + \frac{\gamma x}{(1+\gamma)} \tag{46}$$

Because this model can be viewed as a single chain of microstates, at large times, the flux through each bond between two microstates must be zero. This leads to the following two conditions on stationary probabilities

$$xP_n^{(1)} = (n+1)P_{n+1}^{(1)} \tag{47}$$

and

$$\gamma P_0^{(1)} = P_0^{(2)} \tag{48}$$

One can see that in eqs 45 the first condition (eq 47) is always satisfied whereas the second condition (eq 48) is generally not valid. This suggests that the analysis of the Poisson with zero spike model should be revisited.

Using the conditions presented in eqs 47 and 48 and combining them with the normalization, $\sum_{n=0}^{\infty} P_n^{(1)} + P_0^{(2)} = 1$, we obtain the following stationary-state probabilities

$$P_0^{(1)} = \frac{1}{(e^x+\gamma)},\quad P_0^{(2)} = \frac{\gamma}{(e^x+\gamma)},\quad P_{n\geq 1}^{(1)} = \frac{1}{(e^x+\gamma)}\frac{x^n}{n!} \tag{49}$$

The distribution is illustrated in Figure 2a. Then, the expressions for the mean number of the produced RNA molecules and the Fano factor are given by

$$\langle n\rangle = \frac{x\,e^x}{(e^x+\gamma)},\quad F = 1 + \frac{\gamma x}{(e^x+\gamma)} \tag{50}$$

One can see that these results differ from the formulas obtained in ref 13 (see eq 46) but they approach each other if very large degradation rates are assumed, i.e., when $\beta \gg \alpha$, and we have $x \approx 0$. However, on the basis of the experimental observations,[6,10] this seems to be unrealistic since it would imply a very low mean number of RNA molecules in the system. On the basis of these results, we can conclude that it is not reasonable to utilize this model for analyzing transcriptional bursting phenomena.

Three-State Model. Since experiments indicate that more than two biochemical states might be involved in transcription,[7] we analyze a three-state model shown in Figure 1e. This will also explicitly illustrate our general multistate kinetic approach. We again employ $P_n^{(i)}(t)$ ($i = 1,2,3$) as the probability to find the system with n RNA transcripts at time t, which are governed by the corresponding master equations. Assuming the system is in a stationary state, we introduce three generating functions

$$g_i(z) = \sum_{n=0}^{\infty} z^n P_n^{(i)} \tag{51}$$

for $i = 1-3$.

Our general method suggests that generating functions should be expanded around $y = 0$, where $y = z - 1$, as given in eq 20. Then, the relevant dynamic properties of the system (mean and variance in the number of produced RNA molecules) can be expressed in terms of the expansion coefficients d_i, f_i, and h_i. From eq 26, one can obtain

$$d_1 = \frac{1}{(1+\gamma_1+\gamma_1\gamma_2)},\quad d_2 = \frac{\gamma_1}{(1+\gamma_1+\gamma_1\gamma_2)},$$
$$d_3 = \frac{\gamma_1\gamma_2}{(1+\gamma_1+\gamma_1\gamma_2)} \tag{52}$$

Then, the average number of RNA transcripts in the three-state model is given by

$$\langle n\rangle = \frac{x_1 + x_2\gamma_1 + x_3\gamma_2}{1+\gamma_1+\gamma_1\gamma_2} \tag{53}$$

To evaluate the second moment, we need to know f_1, f_2, and f_3, which can be obtained by solving the system of eqs 17−19. It can be presented conveniently in the matrix form

$$\begin{bmatrix} \beta + k_{off}^{(1)} & -k_{on}^{(2)} & 0 \\ -k_{off}^{(1)} & \beta + k_{off}^{(2)} + k_{on}^{(2)} & -k_{on}^{(3)} \\ 0 & -k_{off}^{(2)} & \beta + k_{on}^{(3)} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \\ f_3 \end{bmatrix} = \begin{bmatrix} \alpha_1 d_1 \\ \alpha_2 d_2 \\ \alpha_3 d_3 \end{bmatrix} \tag{54}$$

Solving this system leads to the following expressions for coefficients f_i

$$f_1 = [x_1(\beta^2 + \beta k_{off}^{(2)} + \beta k_{on}^{(2)} + \beta k_{on}^{(3)} + k_{on}^{(2)}k_{on}^{(3)}) + x_2 k_{off}^{(1)}(\beta + k_{on}^{(3)}) + x_3 k_{off}^{(1)}k_{off}^{(2)}]/YS \tag{55}$$

DOI: 10.1021/acs.jpcb.8b09676
J. Phys. Chem. B 2018, 122, 11969−11977

$$f_2 = [x_2\gamma_1(\beta^2 + \beta k_{off}^{(1)} + \beta k_{on}^{(3)} + \beta k_{on}^{(3)} + k_{off}^{(1)}k_{on}^{(3)}) + x_1 k_{off}^{(1)}(\beta + k_{on}^{(3)}) + x_3\gamma_1 k_{off}^{(2)}(\beta + k_{off}^{(1)})]/YS \quad (56)$$

$$f_3 = [x_3\gamma_1\gamma_2(\beta^2 + \beta k_{off}^{(1)} + \beta k_{off}^{(2)} + \beta k_{on}^{(2)} + k_{off}^{(1)}k_{off}^{(2)}) + x_2\gamma_1 k_{off}^{(2)}(\beta + k_{off}^{(1)}) + x_1 k_{off}^{(1)}k_{off}^{(2)}]/YS \quad (57)$$

where the auxiliary functions Y and S are defined as

$$Y = 1 + \gamma_1 + \gamma_1\gamma_2,\ S = (\beta + k_{off}^{(1)} + k_{on}^{(2)})(\beta + k_{off}^{(2)} + k_{on}^{(3)}) - k_{on}^{(2)}k_{off}^{(2)} \quad (58)$$

Finally, utilizing eqs 24 and 25, we derive the explicit formula for the Fano factor in the three-state model

$$F = 1 - \langle n\rangle + [x_1^2(\beta^2 + \beta k_{off}^{(2)} + \beta k_{on}^{(2)} + \beta k_{on}^{(3)} + k_{off}^{(2)}k_{on}^{(3)}) + 2x_1x_2(\beta k_{off}^{(1)} + k_{off}^{(1)}k_{on}^{(3)}) + 2x_1x_3k_{off}^{(1)}k_{off}^{(2)} + 2x_2x_3\gamma_1(\beta k_{off}^{(2)} + k_{off}^{(1)}k_{off}^{(2)}) + x_2^2\gamma_1(\beta^2 + \beta k_{off}^{(1)} + \beta k_{on}^{(3)} + k_{off}^{(1)}k_{on}^{(3)}) + x_3^2\gamma_1\gamma_2(\beta^2 + \beta k_{off}^{(1)} + \beta k_{off}^{(2)} + k_{off}^{(1)}k_{off}^{(2)} + \beta k_{on}^{(2)})]/[\langle n\rangle YS] \quad (59)$$

Our theoretical calculations are fully verified using extensive Monte-Carlo computer simulations with a stochastic Gillespie algorithm.[28] Using analytical results and computer simulations, we investigate different dynamic behaviors in the system. Assuming that the corresponding transition rates between states are the same ($k_{on}^{(i)} = k_{on}$, $k_{off}^{(i)} = k_{off}$), we develop a comprehensive dynamic phase diagram for the three-state model of transcription. This is presented in Figure 5f where the Fano factor is plotted as a function of the normalized transition rates between the states using a contour map. Five different dynamic regimes are identified in the three-state model on the basis of the number of distribution peaks that are associated with the number of independent biochemical states (see Figure 5a–e). We explored more general conditions ($k_{on}^{(i)} \neq k_{on}$, $k_{off}^{(i)} \neq k_{off}$, and other arbitrary transition rates), and similar features and dynamic phase diagrams are observed in all situations.

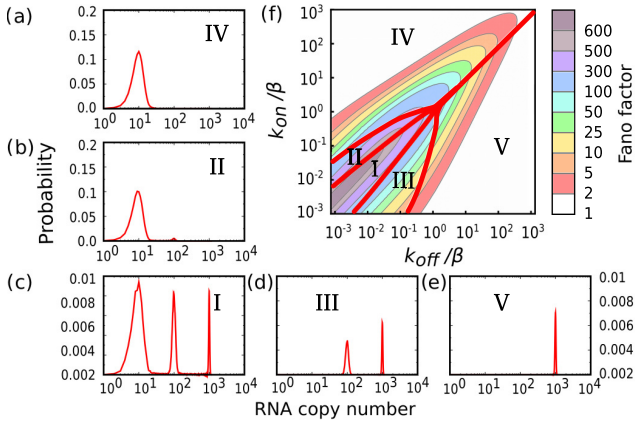


Figure 5. Dynamic properties of RNA production in the three-state model. (a–e) Examples of the five different types of RNA distributions calculated with $\alpha_1 = 10$, $\alpha_2 = 100$, $\alpha_3 = 1000$, and $\beta = 1$: (a) regime IV with $k_{on} = 10$, $k_{off} = 0.1$; (b) regime II with $k_{on} = 0.1$, $k_{off} = 0.01$; (c) regime III with $k_{on} = 0.03$, $k_{off} = 0.1$; (d) regime IV with $k_{on} = 5$, $k_{off} = 0.009$; (e) regime V with $k_{on} = 0.01$, $k_{off} = 10$. (f) A dynamic phase diagram, which shows in the contour plot the amplitude of the Fano factor as the function of the normalized transitions rates between the state. Solid lines give qualitative borders between different types of RNA distributions presented in parts (a)–(e). The rates are in arbitrary units.

To explain this dynamic behavior, we extended theoretical arguments presented in ref 12. In regime I, there are three peaks corresponding to having three independent states. The system experiences the largest fluctuations here because all switching rates are very small, and all three states are explored, i.e., each biochemical state is visited for long periods of time. The Fano factor is maximal in this regime. In regime II, the k_{on} rate increases and this lowers the probability of exploring state 3. As a result, the fluctuations decrease because only two independent biochemical states are probed in this case: state 1 and state 2. Further increase in k_{on} leads to the regime IV where only state 1 is occupied most of the time and the Fano factor has the lowest value. If we move from regime I by increasing the k_{off} rate, we first go into regime III, where mostly the states 2 and 3 are explored. Increasing the k_{off} rate even more leads to the regime V where only state 3 is visited. These explanations of the dynamic behavior of the three-state model are valid for conditions when there is no equilibrium between different biochemical states. In the case of large transition rates (k_{on}, $k_{off} \gg \alpha_i,\beta$), the system will transform into a single equilibrium state where the effective synthesis and degradation rates are averages of the rates of individual states. This is illustrated in Figure 6 where one can see that there is always a single peak in the distribution but the location of the peak depends on the switching equilibrium constant γ.

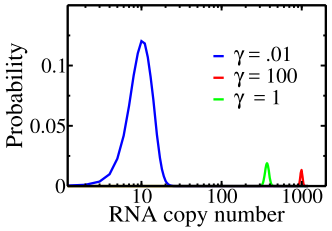


Figure 6. Distributions of the produced RNA molecules for equilibrium conditions for transitions between different biochemical states. The parameter γ is the switching equilibrium constant.

The presented analysis suggests the following physical picture of the transcription processes: In the system with m kinetic states, the number of independent biochemical states can vary from 1 to m depending on the amplitudes of the kinetic rates that couple them. If the transition rates between the states are slow in comparison with growth and degradation rates, the system can explore all m biochemical states and this corresponds to maximal stochastic fluctuations in the system. If transition rates between states are fast in comparison with other rates, then there is an equilibrium between all biochemical states and there is only one independent state (the equilibrium state) in the system. The same picture is observed if only the switching rates to a specific state are high so that other states cannot be visited. This corresponds to the weakest fluctuations in the system. For intermediate values of the transition rates, one expects that equilibrium could be established only between several states or more than one state can be visited and the number of independent biochemical

DOI: 10.1021/acs.jpcb.8b09676
J. Phys. Chem. B 2018, 122, 11969–11977

states can range from 1 to m. This indicates that generally, up to $2m - 1$ different dynamic regimes can be observed and each of them is specified by different numbers and locations of peaks in the distributions of produced RNA molecules. Thus, the underlying chemical-kinetic network of biochemical states that support the transcription might lead to a very rich dynamic behavior, suggesting multiple ways of regulating the transcription processes.

SUMMARY AND CONCLUSIONS

We investigated the phenomenon of transcriptional bursting by developing a general chemical-kinetic approach. It employs a multistate stochastic model, which yields analytical calculations for relevant dynamic properties of transcription. A simple method of evaluating the number of independent biochemical states is proposed and discussed. Our theoretical approach successfully reproduces the analysis for the two-state model of transcription. In addition, we corrected the results for the Poisson with zero spike model, which also has been used to explain transcriptional bursting.[13] We developed a general dynamic phase diagram for a multistate system, and it was specifically illustrated using a three-state model. Our main findings are that the number of peaks in the distribution of RNA molecules correspond to the number of independent biochemical states in the system and for fixed synthesis and degradation rates the degree of stochastic fluctuations directly correlates with this number. By changing the transition rates between the states, the number of independent biochemical states can be varied. The theoretical approach also allows us to evaluate the number of independent biochemical states from the experimental data. These results, which were checked by extensive Monte-Carlo simulations, are able to explain available experimental observations.

Although our theoretical method was able to clarify some features of transcriptional bursting phenomena, it is important to note that our description is rather simplistic because many complex processes that are involved in transcription are not taken into account.[23] For example, it is assumed that all chemical rates are constant in our theoretical approach, whereas it is more realistic to expect some fluctuations due to the variability of the substrate and enzyme concentrations in real biological cells. Rate fluctuations may also be caused by transcriptional feedback controls found in some bacterial systems.[26,27] Recent experiments also suggest that transcription is strongly affected by interactions between RNA polymerases on DNA, but our method does not take this into account.[16] It is still unclear what specific molecular events force switches between different biochemical states in transcription.[5] Another puzzling observation, which again cannot be explained in our approach, is that the transcriptional bursting is mostly gene-independent in bacteria whereas in eukaryotes, it displays more gene-specific behavior, although there is still no full agreement on these results.[5] It is also important to note that the substantial noise in the gene regulation dynamics can also be generated in the process of diffusion of transcription factors to their binding sites, particularly for low-concentration transcription factors and for transient dynamics. This has been studied recently in various theoretical models,[29−32] but it is not accounted in our theoretical approach. Furthermore, the presented theoretical approach does not take into account important feedback regulation processes.[26,27] However, despite these shortcomings, our theoretical model provides a simple, clear and, most importantly, quantitative physical−chemical description of the transcriptional bursting phenomena. It will be important to analyze our theoretical results using more advanced theoretical and computational methods, as well as to test them in experiments.

AUTHOR INFORMATION

Corresponding Author

*E-mail: tolya@rice.edu. Phone: +1 713 3485672.

ORCID

Anatoly B. Kolomeisky: 0000-0001-5677-6690

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The work was supported by the Welch Foundation (C-1559), the NSF (CHE-1664218), and the Center for Theoretical Biological Physics sponsored by the NSF (PHY-1427654). We also would like to thank Dr. Ido Golding for critical reading of the manuscript and for many useful comments and suggestions.

REFERENCES

(1) Lodish, H. F. *Molecular Cell Biology*, 6th ed.; W. H. Freeman and Company, 2008.

(2) Jørgensen, M. G.; van Raaphorst, R.; Veening, J.-W. Noise and Stochasticity in Gene Expression: A Pathogenic Fate Determinant. In *Methods in Microbiology*; Harwood, C., Wipat, A., Eds.; Microbial Synthetic Biology; Academic Press, 2013; Vol. *40*, pp 157−175.

(3) Sanchez, A.; Choubey, S.; Kondev, J. Regulation of Noise in Gene Expression. *Annu. Rev. Biophys.* **2013**, *42*, 469−491.

(4) Chen, H.; Larson, D. R. What Have Single-Molecule Studies Taught Us About Gene Expression? *Genes Dev.* **2016**, *30*, 1796−1810.

(5) Lenstra, T. L.; Rodriguez, J.; Chen, H.; Larson, D. R. Transcription Dynamics in Living Cells. *Annu. Rev. Biophys.* **2016**, *45*, 25−47.

(6) Golding, I.; Paulsson, J.; Zawilski, S. M.; Cox, E. C. Real-Time Kinetics of Gene Activity in Individual Bacteria. *Cell* **2005**, *123*, 1025−1036.

(7) Corrigan, A. M.; Tunnacliffe, E.; Cannon, D.; Chubb, J. R. A Continuum Model of Transcriptional Bursting. *eLife* **2016**, *5*, No. e13051.

(8) Featherstone, K.; Hey, K.; Momiji, H.; McNamara, A. V.; Patist, A. L.; Woodburn, J.; Spiller, D. G.; Christian, H. C.; McNeilly, A. S.; Mullins, J. J.; et al. Spatially Coordinated Dynamic Gene Transcription in Living Pituitary Tissue. *eLife* **2016**, *5*, No. e08494.

(9) Peccoud, J.; Ycart, B. Markovian Modeling of Gene-Product Synthesis. *Theor. Popul. Biol.* **1995**, *48*, 222−234.

(10) So, L.-h.; Ghosh, A.; Zong, C.; Sepúlveda, L.; Segev, R.; Golding, I. General Properties of Transcriptional Time Series in *Escherichia coli*. *Nat. Genet.* **2011**, *43*, 554−560.

(11) Paulsson, J. Models of Stochastic Gene Expression. *Phys. Life Rev.* **2005**, *2*, 157−175.

(12) Munsky, B.; Neuert, G.; van Oudenaarden, A. Using Gene Expression Noise to Understand Gene Regulation. *Science* **2012**, *336*, 183−187.

(13) Chong, S.; Chen, C.; Ge, H.; Xie, X. Mechanism of Transcriptional Bursting in Bacteria. *Cell* **2014**, *158*, 314−326.

(14) Zenklusen, D.; Larson, D. R.; Singer, R. H. Single-RNA Counting Reveals Alternative Modes of Gene Expression in Yeast. *Nat. Struct. Mol. Biol.* **2008**, *15*, 1263−1271.

(15) Larson, D. R.; Zenklusen, D.; Wu, B.; Chao, J.; Singer, R. Real-Time Observation of Transcription Initiation and Elongation on an Endogenous Yeast Gene. *Science* **2011**, *332*, 475−478.

(16) Fujita, K.; Iwaki, M.; Yanagida, T. Transcriptional Bursting Is Intrinsically Caused by Interplay Between RNA Polymerases on DNA. *Nat. Commun.* **2016**, *7*, No. 13788.

DOI: 10.1021/acs.jpcb.8b09676
J. Phys. Chem. B 2018, 122, 11969−11977

(17) Rieckh, G.; Tkacik, G. Noise and Information Transmission in Promoters with Multiple Internal States. *Biophys. J.* **2014**, *106*, 1194−1204.
(18) Hey, K. L.; Momiji, H.; Featherstone, K.; Davis, J. R.; White, M. R.; Rand, D. A.; Finkenstädt, B. A Stochastic Transcriptional Switch Model for Single Cell Imaging Data. *Biostatistics* **2015**, *16*, 655−669.
(19) Jenkins, D. J.; Finkenstädt, B.; Rand, D. A. A temporal Switch Model for Estimating Transcriptional Activity in Gene Expression. *Bioinformatics* **2013**, *29*, 1158−1165.
(20) Raj, A.; Oudenaarden, A. Nature, Nurture or Chance: Stochastic Gene Expression and Its Consequences. *Cell* **2008**, *135*, 216−226.
(21) Sevier, S. A.; Kessler, D. A.; Levine, H. Mechanical Bounds to Transcriptional Noise. *Proc. Natl. Acad. Sci. U.S.A.* **2016**, *113*, 13983−13988.
(22) Sevier, S. A.; Levine, H. Mechanical Properties of Transcription. *Phys. Rev. Lett.* **2017**, *118*, No. 268101.
(23) Sevier, S. A.; Levine, H. Properties of Gene Expression and Chromatin Structure With Mechanically Regulated Elongation. *Nucleic Acids Res.* **2018**, *46*, 5924−5934.
(24) Mugler, A.; Walczak, A. M.; Wiggins, C. H. Spectral Solutions to Stochastic Models of Gene Expression With Bursts and Regulation. *Phys. Rev. E* **2009**, *80*, No. 041921.
(25) Zhou, T.; Zhang, J. Analytical Results for a Multistate Gene Model. *SIAM J. Appl. Math.* **2012**, *72*, 789−818.
(26) Kumar, N.; Platini, T.; Kulkarni, R. V. Exact Distributions for Stochastic Gene Expression Models with Bursting and Feedback. *Phys. Rev. Lett.* **2014**, *113*, No. 268105.
(27) Jia, C.; Xie, P.; Chen, M.; Zhang, M. Q. Stochastic fluctuations Can Reveal the Feedback Signs of Gene Regulatory Networks at the Single-molecule Level. *Sci. Rep.* **2017**, *7*, No. 16037.
(28) Gillespie, D. T. Exact Stochastic Simulation of Coupled Chemical Reactions. *J. Phys. Chem.* **1977**, *81*, 2340−2361.
(29) Pulkkinen, O.; Metzler, R. Distance Matters: The Impact of Gene Proximity in Bacterial Gene Regulation. *Phys. Rev. Lett.* **2013**, *110*, No. 198101.
(30) Godec, A.; Metzler, R. First Passage Time Statistics for Two-Channel Diffusion. *J. Phys. A: Math. Theor.* **2017**, *50*, No. 084001.
(31) Pulkkinen, O.; Metzler, R. Variance-Corrected Michaelis-Menten Equation Predicts Transient Rates of Single-Enzyme Reactions and Response Times in Bacterial Gene-Regulation. *Sci. Rep.* **2015**, *5*, No. 17820.
(32) Godec, A.; Metzler, R. Universal Proximity Effect in Target Search Kinetics in the Few-Encounter Limit. *Phys. Rev. X* **2016**, *6*, No. 041037.

DOI: 10.1021/acs.jpcb.8b09676
J. Phys. Chem. B 2018, 122, 11969−11977