PAPER

Constraining the complexity of promoter dynamics using fluctuations in gene expression

IOP Publishing Phys. Biol. 17 (2020) 015001 https://doi.org/10.1088/1478-3975/ab4e57

Physical Biology

PAPER

RECEIVED
20 August 2019

REVISED
3 October 2019

ACCEPTED FOR PUBLICATION
16 October 2019

PUBLISHED
1 November 2019

Constraining the complexity of promoter dynamics using fluctuations in gene expression

Niraj Kumar and Rahul V Kulkarni

Department of Physics, University of Massachusetts Boston, Boston, MA 02125, United States of America

E-mail: niraj.kumar@umb.edu (N K) and rahul.kulkarni@umb.edu (R V K)

Keywords: stochastic gene expression, transcriptional bursting, promoter switching

Supplementary material for this article is available online

Abstract

Gene expression is an inherently stochastic process with transcription of mRNAs often occurring in bursts: short periods of activity followed by typically longer periods of inactivity. While a simple model involving switching between two promoter states has been widely used to analyze transcription dynamics, recent experimental observations have provided evidence for more complex kinetic schemes underlying bursting. Specifically, experiments provide evidence for complexity in promoter dynamics during the switch from the transcriptionally inactive to the transcriptionally active state. An open question in the field is: what is the minimal complexity needed to model promoter dynamics and how can we determine this? Here, we show that measurements of mRNA fluctuations can be used to set fundamental bounds on the complexity of promoter dynamics. We study models wherein the switching time distribution from transcriptionally inactive to active states is described by a general waiting-time distribution. Using approaches from renewal theory and queueing theory, we derive analytical expressions which connect the Fano factor of mRNA distributions to the waiting-time distribution for promoter switching between inactive and active states. The results derived lead to bounds on the minimal number of promoter states and thus allow us to derive bounds on the minimal complexity of promoter dynamics based on single-cell measurements of mRNA levels.

Introduction

Gene expression in single cells is a stochastic process involving multiple biochemical reactions which can potentially give rise to large cell-to-cell variability in the levels of mRNAs/proteins [1–21]. This intrinsic randomness in gene expression can drive phenotypic variations even in an isogenic population and is known to play critical roles in cell-fate decisions and diverse cellular processes [22]. The quantification of molecular mechanisms involved in gene expression is thus an important step in understanding variability in cellular phenotypes and their responses to external perturbations.

An important feature of gene expression in single cells is that it can occur in a sporadic fashion, characterized by synthesis of mRNAs in short bursts followed by typically longer periods of inactivity. In recent years, multiple experimental observations have provided evidence for such bursty synthesis of mRNAs [23–30]. Such transcriptional bursting is known to increase fluctuations in gene expression, and can thus play a significant role in generating phenotypic variability in a clonal population of cells. At the simplest level, the emergence of such bursting behavior can be understood by analyzing a two-state promoter model, also known as the standard model of gene expression [7, 31–33]. The standard model of gene expression posits that a promoter can exist in either a transcriptionally inactive state (*off* state) or transcriptionally active state (*on* state) with constant rates of switching between promoter states. However, transcription is a complex process involving multiple rate limiting steps as reported in several theoretical and experimental studies [34–40]. For example, recent experimental observations on mammalian cells have provided evidence for multiple rate-limiting steps between the transition from the transcriptionally inactive state to active state [7, 41, 42]. However, promoter switching from the *on* state to the *off* state has been reported to

occur with essentially a single rate-limiting step and thus can be modeled by a constant switching rate [43]. An important open problem in the field is estimation of the number of internal promoter states involved in the switching between inactive to active promoter states.

Recent research in nonequilibrium statistical mechanics has led to the derivation of thermodynamic inequality relations which set fundamental bounds on dissipation based on observations of current fluctuations [44–46]. A natural question that arises is: can we derive similar relations for gene expression that can bound the complexity of promoter dynamics based on observations of current fluctuations? While previous studies have developed approaches for inference of parameters of coarse-grained stchastic models of gene expression using single cell measurements [47], obtaining exact bounds on the number of promoter states for general promoter models is still an open problem. In this paper, we consider a general model of gene expression for which we use approaches from renewal theory and queueing theory to derive fundamental bounds on the minimal number of internal promoter states based on observations of mRNA fluctuations.

The paper is organized as follows. In the next section, we discuss a model for stochastic gene expression wherein promoter switching from its *off* to *on* state is characterized by a general waiting-time distribution. Next, we consider two cases of the model which lead to bounds on the minimum number of internal states involved in the switching process. First, we consider the case that the number of mRNAs created can be experimentally measured as a function of time in single cells allowing us to directly estimate the rate of mRNA production. The second case corresponds to experimental measurements of steady-state mRNA levels in single cells across the population of cells. For both cases, we derive analytic expressions for the Fano-factor associated with mRNA copy numbers, which are then analyzed to estimate the minimum number of states present in the promoter switching process. We conclude with a summary of the results derived.

Model

The experimental observations discussed above [7, 41, 42] indicate that the waiting-time distribution for promoter switching from *off* to *on* states is, in general, non-exponential. However, we do not have a canonical formula for the non-exponential waiting-time distribution since it depends on the number of internal states for the promoter, which can vary from gene to gene. As experimental progress in single-cell measurements has made it possible to count the number of transcripts in different cells, it is of interest to analyze if these measurements can contribute to modeling the complexities of promoter dynamics. The focus of this paper is to derive results that make use of experimentally accessible measurements to derive bounds on the number of internal states involved in promoter switching from *off* to *on* states.

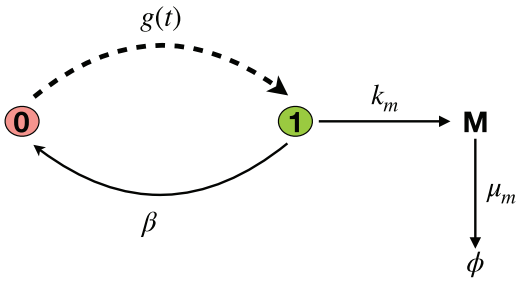

Figure 1. Schematic representation of gene expression model characterized by a general switching time distribution $g(t)$ corresponding to transition of the promoter from an *off* state to the *on* state, denoted by 0 and 1, respectively. Gene in the *on* state either switches back to the *off* state with rate β or synthesizes an mRNA with rate k_m. The mRNA degradation rate is denoted by μ_m.

We consider the model shown in the figure 1 which is based on experimental observations [7, 41, 42] and has been used in previous studies [43]. The model assumes that the promoter has a transcriptionally inactive state (*off* state) and a transcriptionally active state (*on* state). The promoter in the *on* state leads to production of mRNAs with rate k_m, which can then degrade with rate μ_m. The promoter in the *on* state can also switch back to the *off* state with rate β. From the *off* state, the promoter can transition between a general number of internal transcriptionally inactive promoter states before switching back to the *on* state. The waiting-time distribution for switching from the *off* state to the *on* state is represented by $g(t)$, with the notation g emphasizing that it is a general distribution. Note that while the waiting-time distribution is general since the number of promoter states as well as the switching rates between different promoter states can be arbitrary, we are restricting ourselves to models wherein the waiting-time distribution for switching from one promoter state to the next is an exponential distribution.This class of waiting-time distribution is denoted in the queueing literature as phase-type distributions [48]. In order to derive bounds on the number of promoter states, we will use this model for two separate scenarios: first we will consider the case that experimental measurements can access mRNA production in single cells as a function of time. Hence the rate of mRNA production can be directly estimated from available experimental measurements. The second scenario corresponds to the case that we can only access the steady-state distribution of mRNAs across the population of cells.

Inference based on rate of mRNA production

We first consider the case that the mRNA production rate in single cells can be measured experimentally. For this case, the model discussed above (figure 1) will

be used to estimate the number of promoter states involved in the switching of gene from the *off* state to the *on* state. To proceed, we note that, in the model, creation of mRNAs is a renewal process since the time interval between successive arrivals of mRNAs is drawn from the same waiting-time distribution. Correspondingly, the mean number of mRNAs created in time t (in the long-time limit) is given by

$$\langle m(t)\rangle = P_1 k_m t, \tag{1}$$

where P_1 denotes the fraction of time (in the long-time limit) that the promoter is in the transcriptionally active *on* state and $P_1 k_m$ is the average rate of mRNA production. Furthermore, following [49], the variance of the number of mRNAs created in time t is given by:

$$\langle m(t)^2\rangle - \langle m(t)\rangle^2 = \left[\frac{\sigma_\tau^2}{\langle\tau\rangle^2}\right]\langle m(t)\rangle, \tag{2}$$

where $\langle\tau\rangle$ is the mean time interval between successive mRNA arrivals (i.e. renewals), and σ_τ^2 is the variance associated with the corresponding inter-arrival time distribution. It is interesting to note that the Fano factor associated with the number of mRNA arrivals, $F = (\langle m(t)^2\rangle - \langle m(t)\rangle^2)/\langle m(t)\rangle$ is given by the noise associated with the corresponding inter-arrival time distribution, i.e.

$$F = \frac{\sigma_\tau^2}{\langle\tau\rangle^2}. \tag{3}$$

In the following, we will use the Fano-factor associated with mRNA arrivals to derive bounds for the number of states involved in promoter switching from *off* to *on* states.

To proceed further, we note that to derive an expression for F, we need to determine the mean and variance associated with the inter-arrival time distribution ($f(\tau)$) between successive arrivals of mRNAs. As discussed in previous studies [43, 50], an active gene can either produce a mRNA in a single step (i.e. without switching to *off* state) or complete multiple cycles passing through the *off* state before producing an mRNA. Calculations based on this set-up lead to the following explicit expression for the inter-arrival time distribution in the Laplace domain (s) [43]:

$$f_L(s) = \frac{k_m}{k_m + s + (1 - g_L(s))\beta}, \tag{4}$$

with $f_L(s)$ and $g_L(s)$ representing the Laplace transforms of the inter-arrival time distribution for mRNA arrivals and the waiting-time distribution for the promoter to switch from the *off* to the *on* state respectively:

$$f_L(s) = \int_0^\infty f(t)e^{-st}dt,\ g_L(s) = \int_0^\infty g(t)e^{-st}dt. \tag{5}$$

Using equation (5), we can determine the first and second moments of the inter-arrival time distribution ($f(t)$) and the promoter switching-time distribution ($g(t)$) in terms of the coefficients of s and s^2 when $f_L(s)$ and $g_L(s)$ are expanded around $s = 0$, respectively. Using these moments, we obtain the following relation between the noise in the inter-arrival time distribution (η_f) and noise in the promoter switching-time distribution (η_g)

$$\eta_f = 1 + \frac{\beta k_m}{\left(\beta + \frac{1}{\langle t_g\rangle}\right)^2}\left(1 + \eta_g\right). \tag{6}$$

From this equation, we note that noise in the inter-arrival time distribution for mRNAs, and hence the corresponding Fano factor associated with mRNA arrivals (equation (3)), depends on the noise in the the promoter switching time distribution $g(t)$. In particular, if the mean rate of promoter switching ($1/\langle t_g\rangle$) from *off* to *on* states is fixed, the Fano factor of mRNA arrivals is a minimum when the noise in the promoter switching-time distribution is minimized.

Let us now consider the case that the number of transcriptionally inactive states (including the *off* state) is set to a fixed value (n). In this case, the switching-time distribution $g(t)$ can be modeled as a phase-type distribution of order n [48]. For fixed n and fixed mean switching time between *off* and *on* state, we wish to obtain bounds on n based on measurements of mRNA fluctuations. We note that it has been shown that out of all possible phase-type distributions (with fixed mean) of order n, the gamma distribution is the one with the minimum variance [51, 52]. Furthermore, variance of the gamma distribution decreases monotonically with n. This is helpful since it implies that if the variance of the waiting-time distribution for switching between *off* and *on* states (which can be estimated using experimental measurements of mRNA fluctuations using equation (3) above) is less than that of a gamma distribution of order n, then the promoter must have at least $(n + 1)$ promoter states. Using this logic, as elaborated in the following, we can derive general bounds of the mimimal number of promoter states by analyzing gamma waiting-time distributions. Correspondingly, let us consider the case that $g(t)$ is given by a gamma distribution with shape parameter n:

$$g(t;\alpha,n) = \frac{\alpha^n t^{n-1} e^{-\alpha t}}{\Gamma(n)} \tag{7}$$

with α representing the rate associated with a single step, and n denoting the number of steps. This, focusing on the Fano factor for gamma switching-time distributions allows us to set bounds on the Fano factor for mRNA arrivals for *any* switching-time distribution corresponding to fixed n.

For a gamma waiting time distribution with shape parameter n, the exact expression for the Fano-factor (F_n) is given by (see see supplementary material A (stacks.iop.org/PhysBio/17/015001/mmedia))

$$F_n = \frac{\alpha^2 + \beta n\,(n k_m + k_m + \beta n) + 2\alpha\beta n}{(\alpha + \beta n)^2}. \tag{8}$$

In the above formula, F_n is expressed in terms of the model parameters, α, β, n and k_m. However, to gain further insight, it is desirable to express F_n in terms of quantities that are experimentally measurable. In particular, consider the following quantities: the steady-state probability of promoter being in the *on* state P_1, the mean switching rate R of the promoter from its *off* state to the *on* state, and the mean burst size $\langle b \rangle$ i.e. the mean number of mRNAs created in the *on* state just before the promoter switches back to the *off* state. When the switching time distribution is characterized by the gamma distribution, the mean rate of promoter transition from the *off* to *on* state (R) is given by

$$R = \frac{\alpha}{n}. \tag{9}$$

With mean switching rates R and β, the promoter changes stochastically between *on* and *off* states. In the long-time limit, the probability that it is *on* is given by

$$P_1 = \frac{R}{R+\beta}. \tag{10}$$

To quantify bursting, we note that when the gene is in the transcriptionally active *on* state, it can either switch back to the *off* state with rate β or it can produce an mRNA with rate k_m, i.e. the probability that the next reaction for a transcriptionally active promoter is switching to the *off* state is $\beta/(\beta + k_m)$ whereas the probability that the next reaction is production of an mRNA is $k_m/(\beta + k_m)$. Thus the probability that a promoter in the *on* state produces m number of transcripts before it switches to the *off* state is given by the geometric distribution:

$$q(m) = \left[\frac{k_m}{\beta + k_m}\right]^m \frac{\beta}{\beta + k_m}.$$

Thus the mean number of mRNAs produced during the burst is:

$$\langle b \rangle = \sum_{m=0}^{m=\infty} m q(m) = \frac{k_m}{\beta}. \tag{11}$$

The above expressions for R, P_1 and $\langle b \rangle$ will be used for further analysis of the Fano-factor for mRNA arrivals.

Using equations (10) and (11) in equation (8), we derive a simple expression for the Fano-factor of mRNA arrivals (for a gamma waiting time distribution) in terms of the mean burst size ($\langle b \rangle$) and the probability of the promoter being in a transcriptionally inactive state ($P_0 = 1 - P_1$):

$$F_n = 1 + \left[\frac{n+1}{n}\right] P_0^2 \langle b \rangle. \tag{12}$$

Note that we are considering the case of fixed mean switching rates β and R, which implies that P_0 is fixed as well. Thus, the above expression shows that the Fano-factor decreases monotonically with n with a maximum at $n = 1$, i.e. $F_\infty < F_n \leqslant F_1$. Furthermore, since $F_\infty = (1/2)(F_1 + 1)$, we can specify both the lower and upper bounds just in terms of F_1, i.e.

$$\frac{1}{2}(F_1 + 1) < F_n \leqslant F_1. \tag{13}$$

It is convenient to group together the experimentally measurable quantities and to rewrite equation (12) as follows:

$$\mathcal{F}_n = \frac{F_n - 1}{P_0^2 \langle b \rangle} = 1 + \frac{1}{n}, \tag{14}$$

which implies that the preceding inequality can be rewritten as $1 < \mathcal{F}_n \leqslant 2$.

The significance of the above result is the following. Let us consider the experimental analog of $\mathcal{F}_n$ (call it $\mathcal{F}_{ex}$), which has the experimental Fano factor F_{ex} replacing F_n in the preceding equation. To set bounds on the number of promoter states, we need to determine the largest integer n value such that $\mathcal{F}_{ex} < \mathcal{F}_n$. Since the gamma distribution of order n has the lowest variance among all phase-type distributions of order n, and furthermore since F_n decreases monotonically with n, this implies that the promoter must have at least $(n + 1)$ transcriptionally inactive states if $\mathcal{F}_{ex} < \mathcal{F}_n$. Thus measurements of the Fano factor for mRNA arrivals can now be used to set bounds on the minimum number of promoter states involved in the transition from *off* to *on* states.

Inference based on steady-state mRNA levels

The preceding analysis was based on measurements of the rate of mRNA production. We now consider the case that experiments can only access the steady-state distribution of mRNAs. A natural question that arises is: can we use steady-state measurements of mRNA levels to infer similar bounds on the number of promoter states?

For the analysis based on the steady-state measurements, we derive analytical expressions for the steady-state mRNA moments for the model shown in figure 1. Consider first the mapping of gene expression process to models studied in the queuing theory as outlined in previous work [43, 53, 54]. In this mapping, mRNAs are the analogs of customers in a queue: the production of mRNAs is analogous to arrival of customers and degradation of mRNAs is analogous to customers leaving the queue after receiving service. The waiting-time distribution for mRNA degradation is analogous to the customers service time distribution in queuing models. We will assume that distribution of degradation times for mRNAs is exponential. Furthermore, since each mRNAs degrades independently in the model, there are effectively an infinite number of servers in the corresponding queuing model. Correspondingly, the gene expression model as shown in figure 1 is a special case of the $GI/M/\infty$ system in queueing theory: GI stands for a general waiting-time distribution for the inter-arrival times, M denotes Markovian service time distribution for the customers, and ∞ represents infinite servers in the system.

For $GI/M/\infty$ model, we can derive exact expressions for the steady-state moments for the number of customers in the system [43, 53]. The expressions for the steady-state mean and Fano-factor for mRNA levels are given by (see supplementary material B):

$$\langle m_s\rangle = \frac{k_m}{\mu_m}P_1,\ F_s = \frac{1}{2}\left[1+K_g(\mu_m)\right], \quad (15)$$

where P_1 is the steady-state probability that promoter is in the *on* state and $K_g(\mu_m)$ denotes the gestation-factor which characterizes the arrival process for mRNAs and is given by

$$K_g(\mu_m) = 1+2\left[\frac{f_L(\mu_m)}{1-f_L(\mu_m)} - \frac{P_1 k_m}{\mu_m}\right], \quad (16)$$

with $f_L(s)$ denoting the Laplace transform of the waiting time distribution between successive arrivals of mRNAs.

We note that the gestation-factor $K_g(\mu_m)$ depends on the explicit form of the switching-time distribution ($g(\tau)$), which is a general waiting time distribution. Let us consider a limiting case in which the parameter $\mu_m\langle\tau\rangle \ll 1$, wherein τ is the random variable corresponding to the inter-arrival time for mRNAs, as considered in previous studies [47, 53]. In this limit, K_g can be approximated as $K_g \approx \sigma_\tau^2/\langle\tau\rangle^2$ [47, 53]. Correspondingly, the expression for Fano-factor is given by

$$F_s = \frac{1}{2}\left[1+\frac{\sigma_\tau^2}{\langle\tau\rangle^2}\right] = \frac{1+\eta_f}{2}, \quad (17)$$

with η_f as the noise associated with inter-arrival time distribution of mRNAs ($f(t)$).

Note that this expression for the Fano-factor has been derived for the limiting case $K_g \approx \sigma_\tau^2/\langle\tau\rangle^2$. In this limit, we see that the Fano-factor F_s of mRNA steady-state levels is a minimum when the noise in the inter-arrival time distribution η_f is minimum. As we have discussed in the previous section, for a fixed number of promoter states n, η_f is minimum when the switching-time distribution $g(\tau)$ is a gamma distribution. So once again, we can use the gamma waiting-time distribution to derive bounds based on the experimentally measured Fano factor.

Considering the case that $g(\tau)$ corresponds to the gamma distribution (equation (7)), using equations (10) and (15), we obtain:

$$\langle m_s\rangle = \frac{k_m}{\mu_m}\left(\frac{R}{\beta+R}\right);\ R = \alpha/n, \quad (18)$$

and the steady-state the Fano-factor is given by

$$F_{n_s} = 1+\frac{k_m}{\mu_m}\left[\frac{1}{1+\frac{\beta}{\mu_m}\left(1-\left(1+\frac{\mu_m}{nR}\right)^{-n}\right)} - \frac{R}{R+\beta}\right]. \quad (19)$$

The above expression shows that Fano-factor decreases monotonically with n. Its maximum occurs at $n=1$ which corresponds to the standard two promoter states scenario, and its minimum corresponds to $n\to\infty$. The steady-state Fano factor is bounded between these two limiting cases:

$$F_{\infty_s} < F_{n_s} \leqslant F_{1_s}, \quad (20)$$

with F_{1_s} and F_{∞_s} representing steady-state Fano factors for $n=1$ and $n\to\infty$ cases, respectively. Using the general expression for the Fano factor, we can derive explicit expressions for these two bounds. For $n=1$, we get

$$F_{1_s} = 1+\frac{k_m}{\mu_m}\left[\frac{1}{1+\frac{R}{\mu_m}+\frac{\beta}{\mu_m}}\right]\frac{\beta}{R+\beta}, \quad (21)$$

and for the case $n\to\infty$ (using the fact that for $n\to\infty$, $(1+\frac{\mu_m}{nR})^{-n} \to e^{-\frac{\mu_m}{R}}$), we obtain

$$F_{\infty_s} = 1+\frac{k_m}{\mu_m}\left[\frac{1}{1+\frac{\beta}{\mu_m}\left(1-e^{-\frac{\mu_m}{R}}\right)} - \frac{R}{R+\beta}\right]. \quad (22)$$

Now that we have derived upper and lower bounds for the steady-state Fano factor F_{n_s}, we can use these results to derive bounds on the minimum number of promoter states n using measurable quantities such as $\langle m_s\rangle, F_{n_s}, P_0$ or R. Using the following relations:

$$\frac{\beta}{\mu_m} = \frac{\langle m_s\rangle}{\langle b\rangle P_1}, \frac{k_m}{\mu_m} = \frac{\langle m_s\rangle}{P_1}, \frac{R}{\mu_m} = \frac{\langle m_s\rangle}{\langle b\rangle P_0}, \quad (23)$$

in combination with (19) leads to:

$$F_{n_s} = 1+\frac{\langle m_s\rangle}{P_1}\left[\frac{1}{1+\frac{\langle m_s\rangle}{\langle b\rangle P_1}\left(1-\left(1+\frac{\langle b\rangle P_0}{n\langle m_s\rangle}\right)^{-n}\right)} - P_1\right]. \quad (24)$$

The above result expresses the Fano-factor (for gamma waiting time distributions) in terms of experimentally measurable quantities P_1, $\langle m_s\rangle$, $\langle b\rangle$. Now if we have n inactive promoter states, the experimentally measured steady-state Fano factor has to be greater than F_{n_s}, given that the gamma distribution has the least variance among all phase-type distributions of order n. As in the previous section, we can determine the minimum number of promoter states using the inequality $F_{ex} < F_{n_s}$ (where F_{ex} is experimentally measured value of Fano factor) which implies that the number of promoter states must be greater than n. Correspondingly, we derive the following inequality relating the number of inactive promoter states n to experimentally measurable quantities

$$\left(1+\frac{\langle b\rangle P_0}{n\langle m_s\rangle}\right)^{-n} > 1+\frac{\langle b\rangle P_1}{\langle m_s\rangle} - \frac{\langle b\rangle}{F_{ex}+\langle m_s\rangle-1}. \quad (25)$$

The minimum number of promoter states involved in the transition from *off* to *on* states can now be determined using the largest integer satisfying the above inequality.

Conclusions

In conclusion, we have studied a gene expression model wherein promoter switching from a transcriptionally inactive state to an active state is characterized by a general waiting-time distribution. For this generic model of gene expression, we have derived analytical expressions for moments associated with mRNA levels for two scenarios corresponding to experimental measurements of mRNA production rates and steady-state levels in single cells. In both cases, we have obtained analytical expressions relating the Fano factor for mRNAs to properties derived from the waiting-time distribution for promoter switching. For the mRNA production rates case, the Fano factor of interest is directly related to the noise in the waiting-time distribution for successive mRNA arrivals. For the steady-state measurements case, an additional approximation is needed to connect the Fano factor to the noise in the waiting-time distribution. The results derived can then be used to set bounds on the minimal number of promoter states involved in the switching for inactive to active states, using experimentally measurable quantities in combination with theoretical results for the Fano factor corresponding to gamma waiting-time distributions. The significance of these results is that they elucidate that constraints on minimal complexity (in terms of the number of internal promoter states) needed to accurately model the system based on measurements of fluctuations in mRNA levels. The results derived can thus be used to develop minimal models of gene expression dynamics that are consistent with experimentally observed fluctuations in mRNA production.

Acknowledgments

The authors acknowledge funding support from the NSF through Awards DMS- 1413111 and DMS -1854350, and from the NIH through Grants 3U54CA156734-05S3 (as part of the UMass Boston/ Dana Farber-Harvard Cancer Center U54 partnership).

ORCID iDs

Niraj Kumar https://orcid.org/0000-0002-9200-576X

References

[1] Elowitz M B, Levine A J, Siggia E D and Swain P S 2002 *Science* **297** 1183–6
[2] Kaern M, Elston T C, Blake W J and Collins J J 2005 *Nat. Rev. Genet.* **6** 451
[3] Raser J M and O'shea E K 2005 *Science* **309** 2010–3
[4] Sanchez A, Choubey S and Kondev J 2013 *Annu. Rev. Biophys.* **42** 469–91
[5] Eldar A and Elowitz M B 2010 *Nature* **467** 167
[6] Raj A and van Oudenaarden A 2008 *Cell* **135** 216–26
[7] Larson D R 2011 *Curr. Opin. Genet. Dev.* **21** 591–9
[8] Junker J P and van Oudenaarden A 2014 *Cell* **157** 8–11
[9] Munsky B, Neuert G and Van Oudenaarden A 2012 *Science* **336** 183–7
[10] Golding I 2011 *Annu. Rev. Biophys.* **40** 63–80
[11] Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y and Barkai N 2006 *Nat. Genet.* **38** 636
[12] Newman J R, Ghaemmaghami S, Ihmels J, Breslow D K, Noble M, DeRisi J L and Weissman J S 2006 *Nature* **441** 840
[13] Weinberger L, Voichek Y, Tirosh I, Hornung G, Amit I and Barkai N 2012 *Mol. Cell* **47** 193–202
[14] Kumar N, Platini T and Kulkarni R V 2014 *Phys. Rev. Lett.* **113** 268105
[15] Tsimring L S 2014 *Rep. Prog. Phys.* **77** 026601
[16] Battich N, Stoeger T and Pelkmans L 2015 *Cell* **163** 1596–610
[17] Skinner S O, Xu H, Nagarkar-Jaiswal S, Freire P R, Zwaka T P and Golding I 2016 *Elife* **5** e12175
[18] Symmons O and Raj A 2016 *Mol. Cell* **62** 788–802
[19] Nicolas D, Phillips N E and Naef F 2017 *Mol. BioSyst.* **13** 1280–90
[20] Rodriguez J, Ren G, Day C R, Zhao K, Chow C C and Larson D R 2019 *Cell* **176** 213–26
[21] Mackey M C, Tyran-Kaminska M and Yvinec R 2013 *SIAM J. Appl. Math.* **73** 1830–52
[22] Balázsi G, van Oudenaarden A and Collins J J 2011 *Cell* **144** 910–25
[23] Golding I, Paulsson J, Zawilski S M and Cox E C 2005 *Cell* **123** 1025–36
[24] Chubb J R, Trcek T, Shenoy S M and Singer R H 2006 *Curr. Biol.* **16** 1018–25
[25] Raj A, Peskin C S, Tranchina D, Vargas D Y and Tyagi S 2006 *PLoS Biol.* **4** e309
[26] So L H, Ghosh A, Zong C, Sepúlveda L A, Segev R and Golding I 2011 *Nat. Genet.* **43** 554
[27] Taniguchi Y, Choi P J, Li G W, Chen H, Babu M, Hearn J, Emili A and Xie X S 2010 *Science* **329** 533–8
[28] Zong C, So L H, Sepúlveda L A, Skinner S O and Golding I 2010 *Mol. Syst. Biol.* **6** 440
[29] Ochiai H, Sugawara T, Sakuma T and Yamamoto T 2014 *Sci. Rep.* **4** 7125
[30] Senecal A, Munsky B, Proux F, Ly N, Braye F E, Zimmer C, Mueller F and Darzacq X 2014 *Cell Rep.* **8** 75–83
[31] Peccoud J and Ycart B 1995 *Theor. Population Biol.* **48** 222–34
[32] Shahrezaei V and Swain P S 2008 *Proc. Natl Acad. Sci.* **105** 17256–61
[33] Dobrzyński M and Bruggeman F J 2009 *Proc. Natl Acad. Sci.* **106** 2583–8
[34] Zhang J and Zhou T 2019 *Math. Biosci. Eng.* **16** 6134
[35] Qiu H, Zhang B and Zhou T 2019 *Phys. Rev.* E **100** 012128
[36] Zoller B, Nicolas D, Molina N and Naef F 2015 *Mol. Syst. Biol.* **11** 823
[37] Daigle B J Jr, Soltani M, Petzold L R and Singh A 2015 *Bioinformatics* **31** 1428–35
[38] Coulon A, Chow C C, Singer R H and Larson D R 2013 *Nat. Rev. Genet.* **14** 572
[39] Choubey S, Kondev J and Sanchez A 2018 *Biophys. J.* **114** 2072–82
[40] Choubey S 2018 *Phys. Rev.* E **97** 022402
[41] Suter D M, Molina N, Gatfield D, Schneider K, Schibler U and Naef F 2011 *Science* **332** 472–4
[42] Harper C V *et al* 2011 *PLoS Biol.* **9** e1000607
[43] Kumar N, Singh A and Kulkarni R V 2015 *PLoS Comput. Biol.* **11** e1004292
[44] Barato A C and Seifert U 2015 *Phys. Rev. Lett.* **114** 158101
[45] Horowitz J M and Gingrich T R 2017 *Phys. Rev.* E **96** 020103
[46] Proesmans K and Van den Broeck C 2017 *Europhys. Lett.* **119** 20001
[47] Pedraza J M and Paulsson J 2008 *Science* **319** 339–43
[48] Ross S M 2006 *Introduction to Probability Models* 9th edn (Orlando, FL: Academic)
[49] Nebres P 2011 Renewal theory and its applications (unpublished)
[50] Schnitzer M J and Block S 1995 Statistical kinetics of processive enzymes *Cold Spring Harbor Symposia on Quantitative Biology* vol 60 (Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press) pp 793–802
[51] Arnold B C *et al* 2007 *Stat. Sci.* **22** 407–13
[52] David A and Larry S 1987 *Stoch. Models* **3** 467–73
[53] Jia T and Kulkarni R V 2011 *Phys. Rev. Lett.* **106** 058102
[54] Xu X, Kumar N, Krishnan A and Kulkarni R V 2013 Stochastic modeling of dwell-time distributions during transcriptional pausing and initiation *52nd IEEE Conf. on Decision and Control* (IEEE) pp 4068–73