Journal of Theoretical Biology 465 (2019) 56–62

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/jtb

Quantifying heterogeneity of stochastic gene expression

Keita Iida[a,*], Nobuaki Obata[b], Yoshitaka Kimura[a]

[a] Graduate School of Medicine, Tohoku University, Sendai 980-8575, Japan
[b] Graduate School of Information Sciences, Tohoku University, Sendai 980-8579, Japan

ARTICLE INFO

Article history:
Received 18 May 2018
Revised 27 December 2018
Accepted 3 January 2019
Available online 4 January 2019

Keywords:
Stochastic process
Master equation
Lac operon
Metropolis algorithm

ABSTRACT

The heterogeneity of stochastic gene expression, which refers to the temporal fluctuation in a gene product and its cell-to-cell variation, has attracted considerable interest from biologists, physicists, and mathematicians. The dynamics of protein production and degradation have been modeled as random processes with transition probabilities. However, there is a gap between theory and phenomena, particularly in terms of analytical formulation and parameter estimation. In this study, we propose a theoretical framework in which we present a basic model of a gene regulatory system, derive a steady-state solution, and provide a Bayesian approach for estimating the model parameters from single-cell experimental data. The proposed framework is demonstrated to be applicable for various scales of single-cell experiments at both the mRNA and protein levels and is useful for comparing kinetic parameters across species, genomes, and cell strains.

1. Introduction

Gene expression in prokaryotes and eukaryotes is exposed to various molecular noises. Modern single-cell gene expression analyses have revealed that gene expression levels fluctuate in each cell (Chubb et al., 2006; Elowitz and Leibler, 2000; Golding et al., 2005; Taniguchi et al., 2010) and differ from cell to cell even within a clonal population in an identical environment (Blake et al., 2003; Elowitz et al., 2002; Gardner et al., 2000; Zenklusen et al., 2008). Meanwhile, ordered dynamics—ranging from microscopic to macroscopic levels, such as nearly perfect DNA replications (Shapiro, 2007), cell polarization (Zamparo et al., 2015), and mammalian embryogenesis (Yamanaka et al., 2010)—has also been observed. These discoveries reveal that cells integrate both noisy and accurate molecular processes to make them well organized overall, demonstrating the difference between organisms and machinery. However, there is the question of how cells achieve orchestration through stochastic expression kinetics.

It has been proposed that stochastic expression kinetics is related to its cell-to-cell variation (Elowitz et al., 2002), cellular memory (Acar et al., 2005), cell differentiation (Becskei et al., 2001; Süel et al., 2006), and evolution (Fraser and Kærn, 2009). Accordingly, heterogeneous cellular responses to environmental changes have been thoroughly studied (Losick and Desplan, 2008). For example, the *lac* operon in *Escherichia coli* (Choi et al., 2008; Mettetal et al., 2006; Ozbudak et al., 2004) and *GAL* genes in *Saccharomyces cerevisiae* (Acar et al., 2010; Peng et al., 2015; 2016) produce discriminative unimodal and bimodal distributions of the protein concentration. Another study on population survival suggested that increasing expression noise, rather than the mean expression level, could provide cells with a selective advantage under stress conditions (Blake et al., 2006; Bódi et al., 2017). A theoretical study using single-cell experimental data suggested that the Bicoid/Hunchback system in the early *Drosophila* embryo achieves its optimal control in the sense of maximum information transmission (Tkačik et al., 2008). These observations indicate that cells utilize the protein distribution to adapt to environmental changes. Nevertheless, many previous studies have focused on the mean and variance, not on the benefits gained from the distribution.

To address the aforementioned issue, theoreticians have developed analytical procedures to derive mRNA and protein distributions (Assaf et al., 2011; Bokes et al., 2012; Feng et al., 2012; Friedman et al., 2006; Hornos et al., 2005; Paulsson and Ehrenberg, 2000; Peccoud and Ycart, 1995; Pendar et al., 2013; Schwabe et al., 2012; Shahrezaei and Swain, 2008; Shibata, 2003; Vandecan and Blossey, 2013; Zhang et al., 2013; Zhang and Zhou, 2014; Zhou and Liu, 2015). The dynamics of protein production and degradation have been modeled by a discrete stochastic model (Munsky et al. (2012)), whereas the dynamics of transition probability have been modeled by a master equation. However, the connection between stochastic processes and probability distribu-

* Corresponding author. Institute of Industrial Science, The University of Tokyo, Tokyo 153-8505, Japan
E-mail addresses: kiida1984@med.tohoku.ac.jp, kiida@sat.t.u-tokyo.ac.jp (K. Iida), obata@tohoku.ac.jp (N. Obata), ykimura@med.tohoku.ac.jp (Y. Kimura).

https://doi.org/10.1016/j.jtbi.2019.01.003
0022-5193/

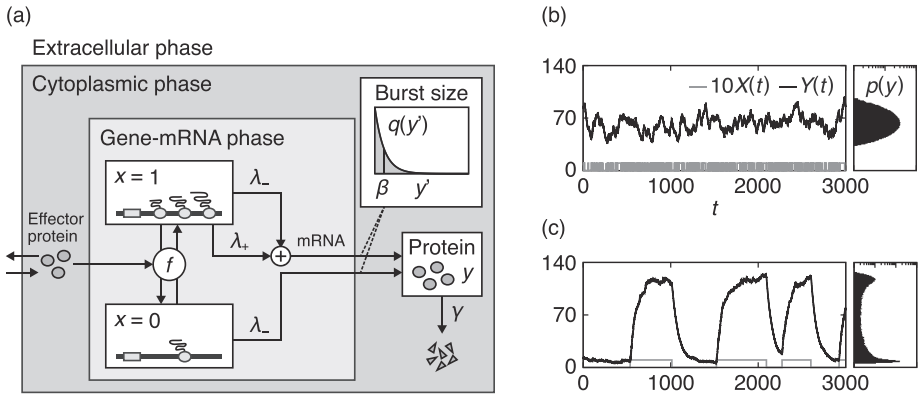

Fig. 1. (a) Schematic representation of the simplified gene regulatory system. (b) and (c) (left) Typical sample paths of $10X(t)$ (gray line) and $Y(t)$ (black line), and (right) $p(y)$, i.e., the probability density of $Y(t)$, plotted in log10 scale. The common parameters are biologically relevant: $\lambda_- = 0.7$ (this value is 10 times larger than the presumable value), $\lambda_+ = 9.73$, $\beta = 0.16$, and $\gamma = 0.014$; and the control parameters (b) $k_{on} = k_{off} = 10\gamma$, and (c) $k_{on} = \gamma/10$ and $k_{off} = 3\gamma/10$.

tion is less studied. This point indicates a gap between theory and phenomena because the master equation does not always represent the stochastic dynamics. Moreover, in practice, many previous studies are based on a discrete model that counts mRNA and protein copy numbers because they are discrete in nature. Meanwhile, in most experiments, single-cell gene expression levels are indirectly observed through the measurement of fluorescence intensity, indicating that a continuous model is needed.

In the present article, we begin with the biophysical modeling of a simple gene regulatory system. First, we formulate the stochastic process of protein production and degradation coupled with an active-or-inactive genetic switch. Second, we introduce a system of master equations and derive an important steady-state solution expressing protein distribution. We show that the solution can be fit to experimental data with an arbitrary measurement scale. Finally, we apply the proposed theory to the thiomethylgalactoside (TMG)-induced system of *lacZYA* expression in *E. coli*, estimate the model parameters from published experimental data (Ozbudak et al., 2004) with Markov chain Monte Carlo (MCMC) methods, and investigate the heterogeneous responses of the *lac* genes to extracellular TMG concentrations. Consequently, the results of this study demonstrate that the proposed theoretical framework is widely applicable to various types of single-cell experiments at both the mRNA and protein levels, such as reporter assays (Bakstad et al., 2012; Coulon et al., 2013; Suter et al., 2011), the MS2-GFP system (Chubb et al., 2006; Golding et al., 2005), fluorescent in situ hybridization (Bakstad et al., 2012; Zenklusen et al., 2008), and flow cytometry (Bódi et al., 2017; Peng et al., 2016).

2. Methods

2.1. Model

Starting with a simplified gene regulatory system, as shown in Fig. 1(a), we assume the following three phases: the extracellular phase is a molecular reservoir in which concentrations remain constant; the cytoplasmic phase contains a vast amount of protein molecules, and the system size is large enough to keep the condition dilute and well stirred (to benefit from the deterministic chemical reaction rate equation (Wallace et al., 2012)); and the gene-mRNA phase is a small-size subsystem in which nonnegligible molecular noises exist. We also assume that the system consists of two variables, x and y, and a gene-protein interaction, f, where $x = 1$ and $x = 0$ denote the active and inactive states of the gene, respectively, and y is the amount of the synthesized protein Y. For the gene-protein interaction, we assume that the probability of x switching between 0 and 1 per unit time dt is described by $x + f(x)\mathrm{d}t$, where

$$f(x) = k_{on}(1 - x) - k_{off}x. \tag{1}$$

Here, k_{on} and k_{off} are the transition rates, depending on the concentrations of effector proteins. Assume that the influx and efflux of the effectors through the cytoplasmic membrane are always in equilibrium (Choi et al., 2008; Yagil and Yagil, 1971), and let k_{on} and k_{off} always be constant. We note that the scheme in Fig. 1(a) is inspired by the well-known two-state model discussed in Friedman et al. (2006); Peccoud and Ycart (1995); Shahrezaei and Swain (2008). In a sense, our model is a continuous version of those proposed in Peccoud and Ycart (1995); Shahrezaei and Swain (2008). The greatest benefit of our approach is that our model has a scalability and continuity with respect to the protein level, which will be discussed later.

Based on the concept of fast-slow dynamics (Popović et al., 2016; Shahrezaei and Swain, 2008), we assume that the unit time is considerably larger and smaller than the lifetimes of the mRNA and protein, respectively, and that Ys are produced through a bursting manner from their mRNA in the genetic phase, whereas they are degraded in a continuous manner in the cytoplasmic phase. We also define the following: λ_- and λ_+ are the burst frequencies, β is the mean burst size per burst, and γ is the decay constant. Here, λ_- represents the basal transcription rate from the inactive promoter, which is referred to as "promoter leakage" (Zhou and Liu, 2015). Based on these assumptions, we formulate the time evolutions of $X \in \{0, 1\}$ and $Y \in \mathbb{R}_{>0}$ by the following mixed random process:

$$X(t + \mathrm{d}t) \sim \mathrm{Ber}(X(t) + f(X(t))\mathrm{d}t), \tag{2}$$

$$\mathrm{d}Y(t) = -\gamma Y(t)\mathrm{d}t + \mathrm{dC}_-(t; \lambda_-, \beta) + X(t)\mathrm{dC}_+(t; \lambda_+, \beta), \tag{3}$$

where "$\sim$" denotes random sampling, and $\mathrm{Ber}(p)$ is the Bernoulli distribution, whose random variable takes a value of 1 with the probability of p; otherwise, it is 0. Here, $\mathrm{dC}_i(t)$ $(i \in \{+, -\})$ is the compound Poisson white noise (or shot noise) (Denisov et al., 2009; Pirrotta, 2007) defined by

$$\mathrm{C}_i(t; \lambda_i, \beta) = \sum_{j=1}^{M_i(t)} R_j U(t - t_j), \tag{4}$$

where $\{M_i(t)\}$ denotes a homogeneous Poisson counting process with the occurrence rate λ_i, $U(t)$ is the unit step function, t_j is the jth arrival time, and $\{R_j\}$ is a sequence of independent identically distributed random burst sizes with a mean of β. Based on experimental observations (Yu et al., 2006) and theoretical assumptions

(Friedman et al., 2006; Paulsson and Ehrenberg, 2000; Shahrezaei and Swain, 2008), we assume that R_j follows an exponential distribution with the assigned probability density $\overline{q}(r) = (1/\beta)e^{-r/\beta}$. The typical sample paths of $X(t)$ and $Y(t)$ for different values of k_{on} and k_{off} are shown in Fig. 1(b) and (c). Note that directly predicting the values of k_{on} and k_{off} may be possible from Fig. 1(c), but such a prediction is difficult from Fig. 1(b). Hence, we need a theoretical procedure to estimate these parameters in any case.

2.2. Analysis

Next, we investigate the probability density functions (PDFs) of $X(t)$ and $Y(t)$. Since $dC_-(t)$ and $dC_+(t)$ are independent random processes, we can analyze the following systems individually:

$$dY_-(t) = -\gamma Y_-(t)dt + dC_-(t;\lambda_-,\beta), \tag{5}$$

$$dY_+(t) = -\gamma Y_+(t)dt + X(t)dC_+(t;\lambda_+,\beta), \tag{6}$$

where $Y(t) = Y_-(t) + Y_+(t)$. Let $p(t, y)$, $p_-(t, y)$, and $p_+(t, x, y)$ be the transition PDFs of $Y(t)$, $Y_-(t)$, and $Y_+(t)$, respectively, and let $p_+(t,y) = \sum_{x=0}^{1} p_+(t,x,y)$. According to the procedures in Ref. Denisov et al. (2009); Pirrotta (2007), we obtain the following system of integro-differential equations along with the normalization condition:

$$\frac{\partial p_-(\tau,y)}{\partial \tau} = \frac{\partial}{\partial y}(y p_-(\tau,y)) + \Lambda_-\left(\int_0^y \overline{q}(y')p_-(\tau,y-y')dy' - p_-(\tau,y)\right), \tag{7}$$

$$\frac{\partial p_+(\tau,0,y)}{\partial \tau} = \frac{\partial}{\partial y}(y p_+(\tau,0,y)) - K_{on}p_+(\tau,0,y) + K_{off}p_+(\tau,1,y), \tag{8}$$

$$\frac{\partial p_+(\tau,1,y)}{\partial \tau} = \frac{\partial}{\partial y}(y p_+(\tau,1,y)) + K_{on}p_+(\tau,0,y) - K_{off}p_+(\tau,1,y) + \Lambda_+\left(\int_0^y \overline{q}(y')p_+(\tau,1,y-y')dy' - p_+(\tau,1,y)\right), \tag{9}$$

$$\int_0^\infty p_i(\tau,y)dy \equiv 1 \quad (i \in \{+,-\}), \tag{10}$$

where $\tau = \gamma t$, $K_{on} = k_{on}/\gamma$, $K_{off} = k_{off}/\gamma$, and $\Lambda_i = \lambda_i/\gamma$ $(i \in \{+,-\})$. Note that Eq. (7) is the so-called Kolmogorov–Feller equation (Gnedenko, 1988) introduced by Friedman and his coworkers as a model of constitutive gene expression (Friedman et al., 2006).

Let us consider the steady-state problem by setting the left-hand sides of Eqs. (7)–(9) to zero. Let $\overline{p}(y)$, $\overline{p}_-(y)$, and $\overline{p}_+(x,y)$ be the PDFs under the steady state, and let $\overline{p}_+(y) = \sum_{x=0}^{1}\overline{p}_+(x,y)$. The Laplace transforms of Eqs. (7)–(9) yield

$$\overline{p}_-(y) = \frac{y^{\Lambda_- - 1}e^{-y/\beta}}{\Gamma(\Lambda_-)\beta^{\Lambda_-}}, \tag{11}$$

$$\overline{p}_+(x,y) = x + (1-2x)C_0\int_0^1 \frac{y^{a-1}e^{-y/(\beta w)}}{\Gamma(a)(\beta w)^a}\cdot\frac{w^{b-1}(1-w)^{c-b}}{B(b,c-b+1)}dw, \tag{12}$$

$$\overline{p}_+(y) = \int_0^1 \frac{y^{a-1}e^{-y/(\beta w)}}{\Gamma(a)(\beta w)^a}\cdot\frac{w^{b-1}(1-w)^{c-b-1}}{B(b,c-b)}dw, \tag{13}$$

$$\overline{p}(y) = \int_0^y \overline{p}_-(y')\overline{p}_+(y-y')dy', \tag{14}$$

where $C_0 = K_{off}/(K_{on}+K_{off})$, $a+b = K_{on}+K_{off}+\Lambda_+$, $ab = K_{on}\Lambda_+$, and $c = K_{on}+K_{off}$ (see Appendix A for the derivation). Here, Eq. (13) is the PDF of the weighted gamma distribution, whose scale parameter βw is averaged over $w \in (0, 1)$ with the assigned beta distribution as a weight function. Assuming that $K_{on}, K_{off} \ll 1$, Eq. (13) can be approximated as follows:

$$\overline{p}_+(y) \approx \int_0^1 \frac{y^{\Lambda_+ - 1}e^{-y/(\beta w)}}{\Gamma(\Lambda_+)(\beta w)^{\Lambda_+}}\cdot\frac{w^{K_{on}-1}(1-w)^{K_{off}-1}}{B(K_{on},K_{off})}dw \quad (K_{on}, K_{off} \ll 1), \tag{15}$$

where Λ_+, K_{on}, and K_{off} serve as the shape parameters of the protein distribution, and β is the scale parameter. In fact, as shown in Eqs. (11)–(13), changing the value of β only results in the scaling of $\overline{p}_i(y)$ $(i \in \{+,-\})$. In other words, for any $k > 0$, the following identity holds:

$$\overline{p}_i(ky;\tilde{\boldsymbol{\theta}}_i)(kdy) = \overline{p}_i(y;\boldsymbol{\theta}_i)dy, \tag{16}$$

where $\tilde{\boldsymbol{\theta}}_i = (k_{on}, k_{off}, \lambda_i, k\beta, \gamma)$, $\boldsymbol{\theta}_i = (k_{on}, k_{off}, \lambda_i, \beta, \gamma)$, and $i \in \{+,-\}$. This result indicates that Eqs. (11)–(14) can be fit to various experimental data with an arbitrary measurement scale of y only by changing the value of β. In addition, we found that the protein distribution defined by Eq. (13) is analogous to the "Poisson-Beta distribution", which is the solution of the discrete model for the birth-and-death process of the mRNA copy number with an all-or-none genetic switch (Iida and Kimura, 2015; Kim and Marioni, 2013; Stinchcombe et al., 2012). Through the analysis, we can also calculate any order of moment (see Appendix B for the calculation) and discuss sources of protein noise (see Appendix C), thereby providing another aspect of our results and a comparison with an earlier report (Shahrezaei and Swain, 2008).

3. Results

Finally, we apply our theory to the TMG-induced system of *lacZYA* expression in *E. coli*, as reported in 2004 (see Figure 2b in Ozbudak et al. (2004)). In the study, cells with well-defined initial states, either uninduced or fully induced, were used, and the fluorescence intensity of the green fluorescent protein (GFP) under the control of *lacZYA* promoter in each cell toward the various TMG concentrations was measured. The remaining definitions and assumptions in our study are as follows: (i) x and y denote the activity of the *lac* promoter and the intracellular levels of the GFP, respectively; (ii) $\overline{p}(y)$ denotes the cell-to-cell variation of y at the steady state; (iii) $\lambda_- + \lambda_+$, β, and γ are the maximum transcription rate defined for the fully induced population, the average scaled number of protein molecules being produced during the *lac* mRNA lifetime, and the decay constant of GFP, respectively; (iv) λ_- is the minimum transcription rate defined for the fully repressed population; and (v) (k_{on}, k_{off}) and ($\lambda_-, \lambda_+, \beta$) are dependent on and independent of the extracellular TMG, respectively. Based on assumptions (i)–(v), we estimate the values of λ_-, λ_+ and β from our bibliographic survey and experimental data with the fully repressed and induced populations (Kennell and Riezman, 1977; Leive and Kollin, 1967; Megerle et al., 2008; Ozbudak et al., 2004; So et al., 2011; Yu et al., 2006). Here, the estimated values are $\lambda_- = 0.07$ min^{-1}, $\lambda_+ = 9.73$ min^{-1}, $\beta = 0.16$, and $\gamma = 0.014$ min^{-1} (see Appendix D for the parameter estimation).

As shown in the previous study, most of the GFP distributions are far from Gaussian, indicating that such distributions are poorly characterized by the mean and variance. Hence, we adopt a Bayesian approach utilizing the Metropolis algorithm (Brooks, 1998; Metropolis et al., 1953) with a noninformative prior and infer the joint posterior distributions of k_{on} and k_{off} for various TMG concentrations from the published single-cell data (Ozbudak et al., 2004) (see Appendix E for the parameter estimation). Since each posterior had a single peak over our searched parameter range, we

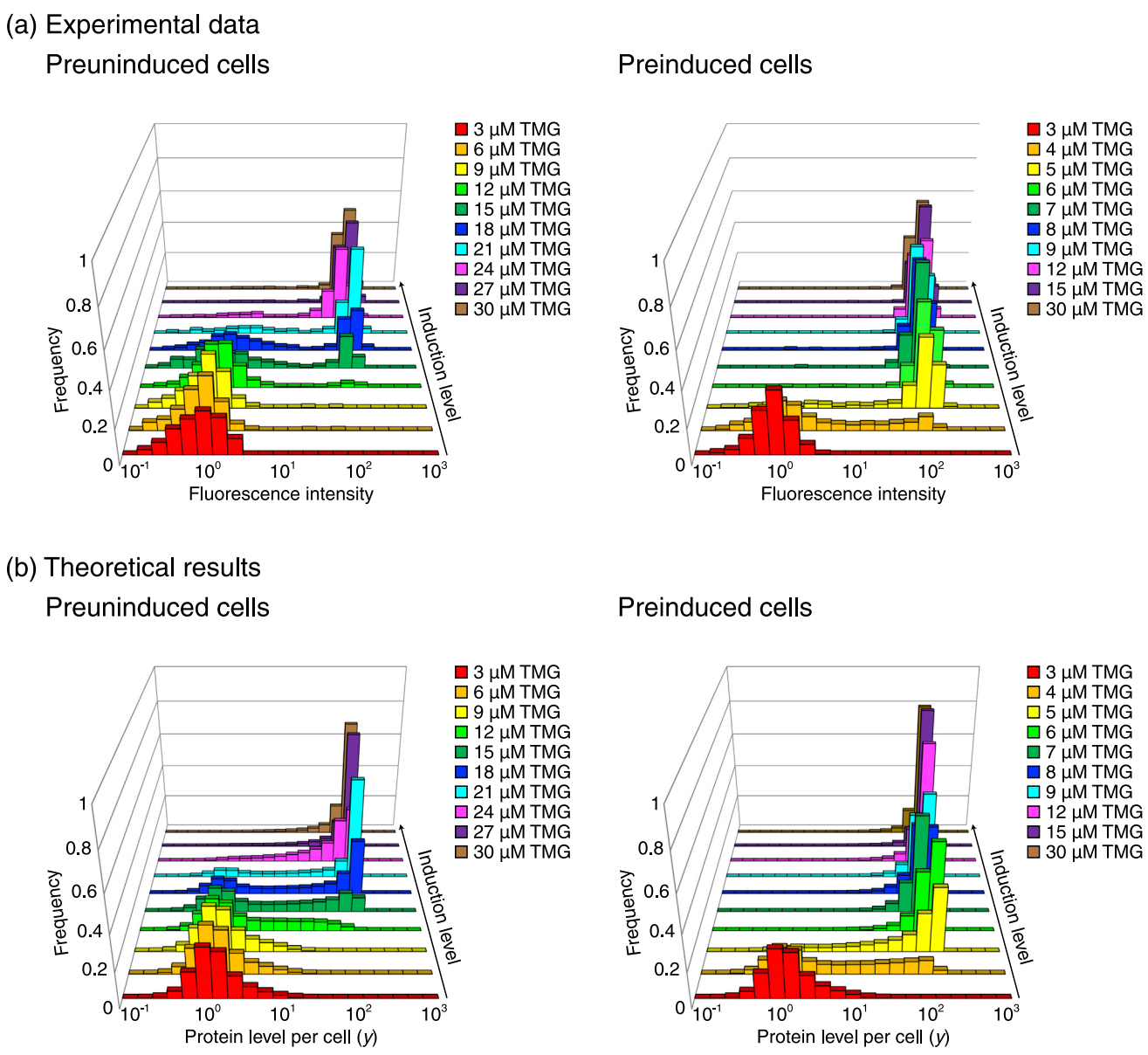

Fig. 2. (color) Histograms obtained from (a) the data Ozbudak et al. (2004) and (b) the independent 10^5 samplings per plot from our model $\overline{p}(y)$ for preuninduced (left panel) and preinduced (right panel) populations. The TMG-independent parameters are $\lambda_- = 0.07$, $\lambda_+ = 9.73$, $\beta = 0.16$, and $\gamma = 0.014$; and TMG-dependent parameters, $k_{\rm on}$ and $k_{\rm off}$, are the mean values obtained from the joint posterior distributions.

inferred the mean values of $k_{\rm on}$ and $k_{\rm off}$. Fig. 2(a) and (b) show histograms obtained from the data (Ozbudak et al., 2004) and the independent samplings from $\overline{p}(y)$ for various TMG concentrations, respectively. As shown in these figures, our parameter estimations using probabilistic model $\overline{p}(y)$ well capture the modality of the experimentally observed distributions across 3–30 μM TMGs with respect to the preuninduced and preinduced populations.

Fig. 3(a) and (b) show the mean values of $k_{\rm on}$ and $k_{\rm off}$ obtained from their posterior distributions as functions of the extracellular TMG levels, and Fig. 3(c) shows the values of $\tau_{\rm on}$ $(=k_{\rm off}^{-1})$ and $\tau_{\rm off}$ $(=k_{\rm on}^{-1})$, which are the average durations of $X(t) = 1$ and $X(t) = 0$, respectively, for both preuninduced and preinduced populations. Consequently, we found that the preuninduced and preinduced populations mainly modulate $k_{\rm off}$ for 3–21 and 3–5 μM TMG, respectively, the values of which are necessary for their half-induction, i.e., the bimodal distribution (see the shaded region in Fig. 3(c)). This result is comparable with those found at the mRNA level (So et al., 2011). In addition, we found that the preuninduced and preinduced cells mainly modulate $k_{\rm on}$ for 24–30 and 6–30 μM TMG, respectively, the values of which are necessary for their further induction.

4. Discussion

We have shown that our theoretical framework can estimate the biological parameters from single-cell data measured as fluorescence intensity. Fig. 3(c) shows that $\tau_{\rm on} \approx 2$ min and $\tau_{\rm off} \approx 345$ min at 3 μM TMG for the preuninduced population. The experimental study with direct measurement using another *E. coli* strain under the control of a repressed *lac* promoter reported that the gene expression burst lasts ~ 3 to 15 min and that the average time between two adjacent expressions is 46 min (Yu et al., 2006). Hence, careful tuning of the parameters λ_-, λ_+, β, and γ in accordance with given experimental data may help researchers accurately quantify the kinetic parameters $k_{\rm on}$ and $k_{\rm off}$. The shape

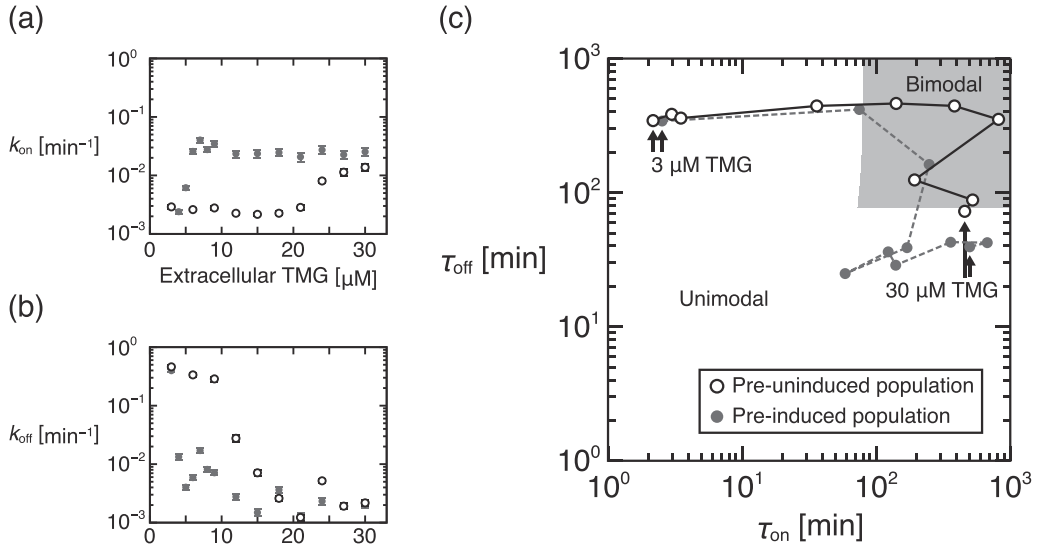

Fig. 3. The inferred mean values of k_{on} and k_{off} for the preuninduced (unfilled circles) and preinduced (filled circles) populations across the various extracellular TMG concentrations. (a) and (b) The rate parameters k_{on} and k_{off} (mean ± standard deviation) versus TMG levels. (c) The average durations of $X(t) = 1$ and $X(t) = 0$ defined by τ_{on} $(= k_{off}^{-1})$ and τ_{off} $(= k_{on}^{-1})$, respectively, for the different TMG concentrations: 3–30 μM with 3 increments for unfilled circles, and 3–9 with 1 increment having 12, 15, and 30 μM for filled circles. The shaded region at the right top corner indicates the bimodality of $\bar{p}(y)$. The solid and dashed lines connecting the points are simply visual guides. The TMG-independent parameters are the same as those for Fig. 2.

of the distribution can also be affected by other factors such as DNA sequence (Saiz, 2012; Sanchez and Golding, 2013), growth rate (Marathe et al., 2012), cell cycle (Soltani and Singh, 2016), partitioning noise (Huh and Paulsson, 2011), and various extrinsic noise Shahrezaei et al., 2008. Incorporating these effects into the model and choosing an effective parameter estimation method are future studies.

Interestingly, Fig. 3(a) and (b) show that k_{on} and k_{off} are nonlinear functions of TMG levels. In this case, if the TMG level is a random variable that obeys a distribution with nonzero variance, then the mean output value of y may deviate from that predicted from deterministic reaction kinetics. This situation occurs at any time when considering a random molecular flux across the cell membrane, a genetic circuit with feedback loops, and so forth. Mathematically, Jensen's inequality may help us predict such deviation provided that k_{on} and k_{off} are convex functions. However, Fig. 3(a) and (b) indicate that k_{on} and k_{off} are nonconvex, which leads to a notorious mixed convex problem (Niculescu and Rovenţa, 2015).

Within the next decade, single-cell experiments with quantitative mathematical biology will enable comparing biologically important parameters such as τ_{on} and τ_{off} across species, genomes, and strains as functions of various environmental conditions, part of which were briefly reviewed by Lionnet and Singer (2012). Gnügge et al. (2016) reported that *E. coli* and *S. cerevisiae* have similar core networks in the lactose and galactose utilizing systems, but the functional role of the network complexity has yet to be determined. Our results show that the transition behaviors of the system can be compared on the (τ_{on}, τ_{off})-plane along with the probability distributions (Figs. 2 and 3(c)), which may provide a better understanding of the quantitative differences across species.

Pioneers have proposed ingenious methods for estimating their model parameters from a finite number of statistics (Dar et al., 2012; Munsky et al., 2009; Zechner et al., 2012), such as expression level, coefficient of variation, and autocorrelation time. They mapped the kinetic features of human gene expression into their parameter space through a genome-wide experiment (Dar et al., 2012). However, their method is based on a finite number of statistics, which can estimate a smaller number of parameters; furthermore, the experiment requires high-resolution real-time monitoring, which is costly. To overcome these issues, a Bayesian or maximum likelihood approach is suitable, as Shahrezaei and Swain (2008) mentioned in their report. In this respect, our framework including MCMC performed well without any concern for the number of statistics.

Choi et al. (2008) experimentally examined various *E. coli* strains with different genetic constructions and found that the binding affinity of the effector protein to its target locus significantly changes the protein distribution. Linking DNA structure, such as looping, chemical modification, and DNA-protein complex formation, to its expression pattern is an important future work.

5. Conclusion

We have proposed a fairly general model of a gene regulatory system along with clear biophysical assumptions, and we formulated this model by both stochastic differential equations (Eqs. (2) and (3)) and corresponding master equations (Eqs. (7)–(10)). We subsequently derived a steady-state solution while avoiding complicated forms such as intricate combinations of hypergeometric functions, which allows one to understand the parameters that affect the shape and scale of the protein distribution (Eq. (15)). We also demonstrated that the solution can be fit to experimental data with an arbitrary measurement scale (Eq. (16)).

As an application, we investigated the TMG-induced system of *lacZYA* expression in *E. coli*. Accordingly, we found that the system mainly modulates k_{off} for the lower or intermediate levels of induction and k_{on} for the higher level, which can be predicted on the (τ_{on}, τ_{off})-plane (Fig. 3(c)). Finally, we conclude that our theoretical framework is widely applicable to various types of single-cell experiments at both the mRNA and protein levels, and it is expected to be useful for predicting the kinetic behaviors of a given biological network, including genetic oscillations (Elowitz and Leibler, 2000; Vilar et al., 2002), cell differentiations (Süel et al., 2006), and evolution (Bódi et al., 2017).

Acknowledgments

The authors thank Satoshi Mizuno, Soichi Ogishima, Naoko Kasahara, Satoshi Nagaie, and Rie Norita for their helpful discus-

sions. This work was partially supported by JSPS KAKENHI Grant number 16H03939.

Supplementary material

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.jtbi.2019.01.003.

References

Acar, D., Becskei, A., van Oudenaarden, A., 2005. Enhancement of cellular memory by reducing stochastic transitions. Nature 435, 228–232. doi:10.1038/nature03524.

Acar, M., Pando, B.F., Arnold, F.H., Elowitz, M.B., van Oudenaarden, A., 2010. A general mechanism for network-dosage compensation in gene circuits. Science 329, 1656–1660. doi:10.1126/science.1190544.

Assaf, M., Roberts, E., Schulten, Z.L., 2011. Determining the stability of genetic switches: explicitly accounting for mrna noise. Phys. Rev. Lett. 106, 248102. doi:10.1103/PhysRevLett.106.248102.

Bakstad, D., Adamson, A., Spiller, D.G., White, M.R.H., 2012. Quantitative measurement of single cell dynamics. Current Opin. Biotech. 23, 103–109. doi:10.1016/j.copbio.2011.11.007.

Becskei, A., Séraphin, B., Serrano, L., 2001. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. EMBO J. 20, 2528–2535. doi:10.1093/emboj/20.10.2528.

Blake, W.J., Balázsi, G., Kohanski, M.A., Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R., Collins, J.J., 2006. Phenotypic consequences of promoter-mediated transcriptional noise. Mol. Cell 24, 853–865. doi:10.1016/j.molcel.2006.11.003.

Blake, W.J., Kærn, M., Cantor, C.R., Collins, J.J., 2003. Noise in eukaryotic gene expression. Nature 422, 633–637. doi:10.1038/nature01546.

Bódi, Z., Farkas, Z., Nevozhay, D., Kalapis, D., Lázár, V., Csörgő, B., Nyerges, Á., Szamecz, B., Fekete, G., Papp, B., Araújo, H., Oliveira, J.L., Moura, G., Santos, M.A.S., Székely Jr, T., Balázsi, G., Pál, C., 2017. Phenotypic heterogeneity promotes adaptive evolution. PLoS Biol. 15, 1–26. doi:10.1371/journal.pbio.2000644.

Bokes, P., King, J.R., Wood, A.T.A., Loose, M., 2012. Exact and approximate distributions of protein and mrna levels in the low-copy regime of gene expression. J. Math. Biol. 64, 829–854. doi:10.1007/s00285-011-0433-5.

Brooks, S.P., 1998. Markov chain monte carlo method and its application. J. R. Stat. Soc. D-Sta. 47, 69–100. doi:10.1111/1467-9884.00117.

Choi, P.J., Cai, L., Frieda, K., Xie, X.S., 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. Science 322, 442–446. doi:10.1126/science.1161427.

Chubb, J.R., Trcek, T., Shenoy, S.M., Singer, R.H., 2006. Transcriptional pulsing of a developmental gene. Current Biol. 16, 1018–1025. doi:10.1016/j.cub.2006.03.092.

Coulon, A., Chow, C.C., Singer, R.H., Larson, D.R., 2013. Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat. Rev. Genet. 14, 572–584. doi:10.1038/nrg3484.

Dar, R.D., Razooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., Weinberger, L.S., 2012. Transcriptional burst frequency and burst size are equally modulated across the human genome. Proc. Natl. Acad. Sci. USA. 109, 17454–17459. doi:10.1073/pnas.1213530109.

Denisov, S.I., Horsthemke, W., Hänggi, P., 2009. Generalized Fokker-planck equation: derivation and exact solutions. Eur. Phys. J. B 68, 567–575. doi:10.1140/epjb/e2009-00126-3.

Elowitz, M.B., Leibler, S., 2000. A synthetic oscillatory network of transcriptional regulators. Nature 403, 335–338. doi:10.1038/35002125.

Elowitz, M.B., Levine, A.J., Siggia, E.D., Swain, P.S., 2002. Stochastic gene expression in a single cell. Science 297, 1183–1186. doi:10.1126/science.1070919.

Feng, H., Hensel, Z., Xiao, J., Wang, J., 2012. Analytical calculation of protein production distributions in models of clustered protein expression. Phys. Rev. E 85, 031904. doi:10.1103/PhysRevE.85.031904.

Fraser, D., Kærn, M., 2009. A chance at survival: gene expression noise and phenotypic diversification strategies. Mol. Microbiol. 71, 1333–1340. doi:10.1111/j.1365-2958.2009.06605.x.

Friedman, N., Cai, L., Xie, X.S., 2006. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Phys. Rev. Lett. 97, 168302. doi:10.1103/PhysRevLett.97.168302.

Gardner, T.S., Cantor, C.R., Collins, J.J., 2000. Construction of a genetic toggle switch in *escherichia coli*. Nature 403, 339–342. doi:10.1038/35002131.

Gnedenko, B.V., 1988. Theory of Probability, 6 ed. Gordon and reach Science Publishers.

Gnügge, R., Dharmarajan, L., Lang, M., Stelling, J., 2016. An orthogonal permease-inducer-repressor feedback loop shows bistability. ACS Synth. Biol. 5, 1098–1107. doi:10.1021/acssynbio.6b00013.

Golding, I., Paulsson, J., Zawilski, S.M., Cox, E.C., 2005. Real-time kinetics of gene activity in individual bacteria. Cell 123, 1025–1036. doi:10.1016/j.cell.2005.09.031.

Hornos, J.E.M., Schultz, D., Innocentini, G.C.P., Wang, J., Walczak, A.M., Onuchic, J.N., Wolynes, P.G., 2005. Self-regulating gene: an exact solution. Phys. Rev. E 72, 051907. doi:10.1103/PhysRevE.72.051907.

Huh, D., Paulsson, J., 2011. Non-genetic heterogeneity from random partitioning at cell division. Nat. Genet. 43, 95–100. doi:10.1038/ng.729.

Iida, K., Kimura, Y., 2015, 1 ed., 1. Springer, New York, pp. 117–120.

Kennell, D., Riezman, H., 1977. Transcription and translation initiation frequencies of the *escherichia coli lac* operon. J. Mol. Biol. 114, 1–21. doi:10.1016/0022-2836(77)90279-0.

Kim, J.K., Marioni, J.C., 2013. Inferring the kinetics of stochastic gene expression from single-cell rna-sequencing data. Genome Biol. 14, 1–12. doi:10.1186/gb-2013-14-1-r7.

Leive, L., Kollin, V., 1967. Synthesis, utilization and degradation of lactose operon mRNA in *escherichia coli*. J. Mol. Biol. 24, 247–269. doi:10.1016/0022-2836(67)90330-0.

Lionnet, T., Singer, R.H., 2012. Transcription goes digital. EMBO Rep. 13, 313–321. doi:10.1038/embor.2012.31.

Losick, R., Desplan, C., 2008. Stochasticity and cell fate. Science 320, 65–68. doi:10.1126/science.1147888.

Marathe, R., Bierbaum, V., Gomez, D., Klumpp, S., 2012. Deterministic and stochastic descriptions of gene expression dynamics. J. Stat. Phys. 148, 608–627. doi:10.1007/s10955-012-0459-0.

Megerle, J., Fritz, G., Gerland, U., Jung, K., Rädler, J.O., 2008. Timing and dynamics of single cell gene expression in the arabinose utilization system. Biophys. J. 95, 2103–2115. doi:10.1529/biophysj.107.127191.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H., Teller, E., 1953. Equation of state calculations by fast computing machines. J. Chem. Phys. 21, 1087–1092. doi:10.1063/1.1699114.

Mettetal, J.T., Muzzey, D., Pedraza, J.M., Ozbudak, E.M., van Oudenaarden, A., 2006. Predicting stochastic gene expression dynamics in single cells. Proc. Natl. Acad. Sci. USA. 103, 7304–7309. doi:10.1073/pnas.0509874103.

Munsky, B., Neuert, G., van Oudenaarden, A., 2012. Using gene expression noise to understand gene regulation. Science 336, 183–187. doi:10.1126/science.1216379.

Munsky, B., Trinh, B., Khammash, M., 2009. Listening to the noise: random fluctuations reveal gene network parameters. Mol. Syst. Biol. 5, 1–7. doi:10.1038/msb.2009.75.

Niculescu, C.P., Rovenţa, I., 2015. Relative convexity and its applications. Aequat. Math. 89, 1389–1400. doi:10.1007/s00010-014-0319-x.

Ozbudak, E.M., Thattai, M., Lim, H.N., Shraiman, B.I., van Oudenaarden, A., 2004. Multistability in the lactose utilization network of *escherichia coli*. Nature 427, 737–740. doi:10.1038/nature02298.

Paulsson, J., Ehrenberg, M., 2000. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. Phys. Rev. Lett. 84, 5447–5450. doi:10.1103/PhysRevLett.84.5447.

Peccoud, J., Ycart, B., 1995. Markovian modelling of gene product synthesis. Theor. Popul. Biol. 48, 222–234. doi:10.1006/tpbi.1995.1027.

Pendar, H., Platini, T., Kulkarni, R.V., 2013. Exact protein distributions for stochastic models of gene expression using partitioning of poisson processes. Phys. Rev. E 87, 042720. doi:10.1103/PhysRevE.87.042720.

Peng, W., Liu, P., Xue, Y., Acar, M., 2015. Evolution of gene network activity by tuning the strength of negative-feedback regulation. Nat. Commun. 6, 1–9. doi:10.1038/ncomms7226.

Peng, W., Song, R., Acar, M., 2016. Noise reduction facilitated by dosage compensation in gene networks. Nat. Commun. 7, 1–8. doi:10.1038/ncomms12959.

Pirrotta, A., 2007. Multiplicative cases from additive cases: extension of Kolmogorov-feller equation to parametric poisson white noise processes. Probabilist. Eng. Mech. 22, 127–135. doi:10.1016/j.probengmech.2006.08.005.

Popović, N., Marr, C., Swain, P.S., 2016. A geometric analysis of fast-slow models for stochastic gene expression. J. Math. Biol. 72, 87–122. doi:10.1007/s00285-015-0876-1.

Saiz, L., 2012. The physics of protein-dna interaction networks in the control of gene expression. J. Phys. Condens Matter 24, 1–15. doi:10.1088/0953-8984/24/19/193102.

Sanchez, A., Golding, I., 2013. Genetic determinants and cellular constraints in noisy gene expression. Science 342, 1188–1193. doi:10.1126/science.1242975.

Schwabe, A., Rybakova, K.N., Bruggeman, F.J., 2012. Transcription stochasticity of complex gene regulation models. Biophys. J. 103, 1152–1161. doi:10.1016/j.bpj.2012.07.011.

Shahrezaei, V., Ollivier, J.F., Swain, P.S., 2008. Colored extrinsic fluctuations and stochastic gene expression. Mol. Syst. Biol. 4, 1–9. doi:10.1038/msb.2008.31.

Shahrezaei, V., Swain, P.S., 2008. Analytical distributions for stochastic gene express. Proc. Natl. Acad. Sci. USA. 105, 17256–17261. doi:10.1073/pnas.0803850105.

Shapiro, J.A., 2007. Bacteria are small but not stupid: cognition, natural genetic engineering and socio-bacteriology. Stud. Hist. Philos. Biol. Biomed. Sci. 38, 807–819. doi:10.1016/j.shpsc.2007.09.010.

Shibata, T., 2003. Fluctuating reaction rates and their application to problems of gene expression. Phys. Rev. E 67, 061906. doi:10.1103/PhysRevE.67.061906.

So, L., Ghosh, A., Zong, C., Sepúlveda, L.A., Segev, R., Golding, I., 2011. General properties of transcriptional time series in *escherichia coli*. Nat. Genet. 43, 554–560. doi:10.1038/ng.821.

Soltani, M., Singh, A., 2016. Effects of cell-cycle-dependent expression on random fluctuations in protein levels. R. Soc. Open Sci. 3, 160578. doi:10.1098/rsos.160578.

Stinchcombe, A.R., Peskin, C.S., Tranchina, D., 2012. Population density approach for discrete mRNA distributions in generalized switching models for stochastic gene expression. Phys. Rev. E 85, 061919. doi:10.1103/PhysRevE.85.061919.

Süel, G.M., Garcia-Ojalvo, J., Liberman, L.M., Elowitz, M.B., 2006. An excitable gene regulatory circuit induces transient cellular differentiation. Nature 440, 545–550. doi:10.1038/nature04588.

Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., Naef, F., 2011. Mammalian genes are transcribed with widely different bursting kinetics. Science 332, 472–474. doi:10.1126/science.1198817.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., Xie, X.S., 2010. Quantifying *e. coli* proteome and transcriptome with single-molecule sensitivity in single cells. Science 329, 533–538. doi:10.1126/science.1188308.

Tkačik, G., Jr., C.G.C., Bialek, W., 2008. Information flow and optimization in transcriptional regulation. Proc. Natl. Acad. Sci. USA. 105, 12265–12270. doi:10.1073/pnas.0806077105.

Vandecan, Y., Blossey, R., 2013. Self-regulatory gene: an exact solution for the gene gate model. Phys. Rev. E 87, 042705. doi:10.1103/PhysRevE.87.042705.

Vilar, J.M.G., Kueh, H.Y., Barkai, N., Leibler, S., 2002. Mechanisms of noise-resistance in genetic oscillators. Proc. Natl. Acad. Sci. USA. 99, 5988–5992. doi:10.1073/pnas.092133899.

Wallace, E.W.J., Gillespie, D.T., Sanft, K.R., Petzold, L.R., 2012. Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. IET Syst. Biol. 6, 102–115. doi:10.1049/iet-syb.2011.0038.

Yagil, G., Yagil, E., 1971. On the relation between effector concentration and the rate of induced enzyme synthesis. Biophys. J. 11, 11–27. doi:10.1016/S0006-3495(71)86192-1.

Yamanaka, Y., Lanner, F., Rossant, J., 2010. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. Development 137, 715–724. doi:10.1242/dev.043471.

Yu, J., Xiao, J., Ren, X., Lao, K., Xie, X.S., 2006. Probing gene expression in live cells, one protein molecule at a time. Science 311, 1600–1603. doi:10.1126/science.1119623.

Zamparo, M., Chianale, F., Tebaldi, C., Cosentino-Lagomarsino, M., Nicodemi, M., Gamba, A., 2015. Dynamic membrane patterning, signal localization and polarity in living cells. Soft Matter 11, 838–849. doi:10.1039/c4sm02157f.

Zechner, C., Ruess, J., Krenn, P., Pelet, S., Peter, M., Lygeros, J., Koeppl, H., 2012. Moment-based inference predicts bimodality in transient gene expression. Proc. Natl. Acad. Sci. USA. 109, 8340–8345. doi:10.1073/pnas.1200161109.

Zenklusen, D., Larson, D.R., Singer, R.H., 2008. Single-rna counting reveals alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. 15, 1263–1271. doi:10.1038/nsmb.1514.

Zhang, J., Nie, Q., He, M., Zhou, T., 2013. An effective method for computing the noise in biochemical networks. J. Chem. Phys. 138, 084106. doi:10.1063/1.4792444.

Zhang, J., Zhou, T., 2014. Promoter-mediated transcriptional dynamics. Biophys. J. 106, 479–488. doi:10.1016/j.bpj.2013.12.011.

Zhou, T., Liu, T., 2015. Quantitative analysis of gene expression systems. Quant. Biol. 3, 168–181. doi:10.1007/s40484-015-0056-8.