bioRxiv preprint doi: https://doi.org/10.1101/316166; this version posted May 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Manuscript submitted to **Biophysical** *Journal*

Article

Quantifying heterogeneity of stochastic gene expression

K. Iida, N. Obata, and Y. Kimura

bioRxiv preprint doi: https://doi.org/10.1101/316166; this version posted May 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ABSTRACT The heterogeneity of stochastic gene expression, which refers to the temporal fluctuation of the gene product and its cell-to-cell variation, has attracted considerable interest from biologists, physicists, and mathematicians. The dynamics of protein production and degradation were 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 it is useful for comparing kinetic parameters across species, genomes, and cell strains.

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 (1–3) and differ from cell to cell even within a clonal population in an identical environment (4–7). Meanwhile, ordered dynamics—ranging from microscopic to macroscopic levels, such as nearly perfect DNA replications (8), cell polarization (9), and mammalian embryogenesis (10)—have 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, the question of how do cells achieve orchestration through stochastic expression kinetics arises.

It has been proposed that stochastic expression kinetics is related to its cell-to-cell variation (5), cellular memory (11), cell differentiation (12, 13), and evolution (14). Accordingly, heterogeneous cellular responses to environmental changes have been thoroughly studied. For example, the *lac* operon in *Escherichia coli* (15–17) and *GAL* genes in *Saccharomyces cerevisiae* (18–20) 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 (21, 22). These observations indicate that cells utilize the protein distribution to adapt to the 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 (23–34). The dynamics of protein production and degradation have been modeled by a discrete stochastic model, whereas the dynamics of transition probability have been modeled by a master equation. However, the connection between stochastic processes and probability distribution 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, the single-cell gene expressions are indirectly observed through the measurement of fluorescence intensity, thus indicating that we need a continuous model.

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 (15) 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 (35–37), MS2-GFP system (2, 3), fluorescent in situ hybridization (7, 36), and flow cytometry (20, 22).

METHODS

Model

Starting with a simplified gene regulatory system, as shown in Fig. 1(a), we assume the following three phases: phase I, i.e., the extracellular phase, is a molecular reservoir in which concentrations remain constant; phase II, i.e., 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 (38)); and phase III, i.e., the gene-mRNA phase, is a small-size subsystem in which non-negligible molecular noises exist. We further assume that the system consists of three variables, x, y, and z, and a gene-protein interaction, f, where $x = 1$ and $x = 0$ denote the active and inactive states of the gene, respectively,

bioRxiv preprint doi: https://doi.org/10.1101/316166; this version posted May 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

and y and z are the amounts of the synthesized protein Y and a certain effector protein Z, respectively. 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

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

Here, $k_{\mathrm{on}}(z)$ and $k_{\mathrm{off}}(z)$ are the transition rates. Assume that the influx and efflux of Z through the cytoplasmic membrane are always in equilibrium (17, 39), and let z always be constant. Henceforth, let $k_{\mathrm{on}}(z) = k_{\mathrm{on}}$, and let $k_{\mathrm{off}}(z) = k_{\mathrm{off}}$.

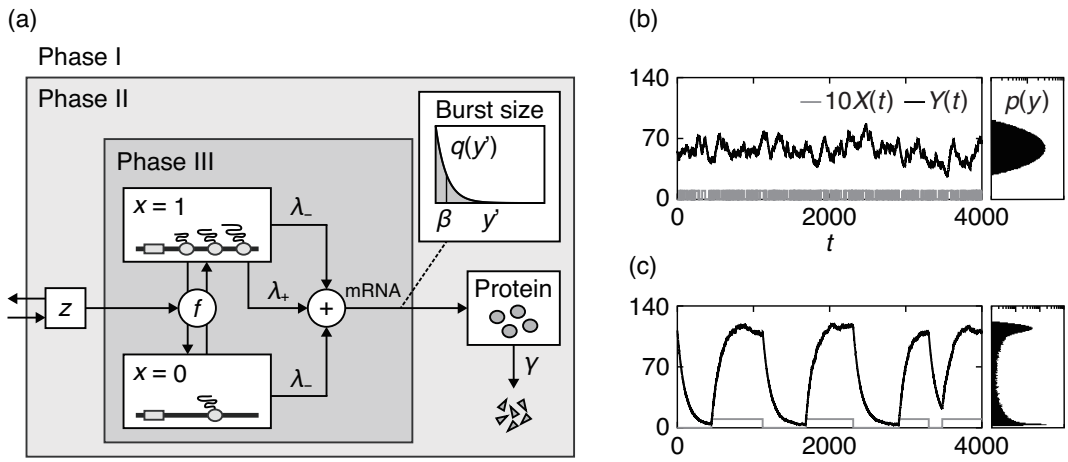

Figure 1: (a) Schematic representation of the simplified gene regulatory system. (b) and (c) (left) Typical sample paths of $10X(t)$ (gray line) and $Y(t)$ (black line), and (right) $p(y)$, i.e., the probability density of $Y(t)$, plotted in log10 scale. The common parameters are biologically relevant: $\lambda_- = 0.5$ (this is slightly larger than the presumable value), $\lambda_+ = 15$, $\beta = 0.074$, and $\gamma = 0.01$; and the control parameters (b) $k_{\mathrm{on}} = k_{\mathrm{off}} = 10\gamma$, and (c) $k_{\mathrm{on}} = k_{\mathrm{off}} = \gamma/10$.

Based on the concept of fast-slow dynamics (27, 40), we assume that the unit time is considerably larger and smaller than the lifetimes of the mRNA and protein, respectively, and 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" (34). 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 1 with the probability of p; otherwise, it is 0. Here, $\mathrm{dC}_i(t)$ ($i \in \{+, -\}$) is the compound Poisson white noise (41, 42) 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 β. Based on experimental observations (43) and theoretical assumptions (24, 26, 27), we assume that R_j follows an exponential distribution with the assigned probability density $\overline{q}(r) = (1/\beta)\mathrm{e}^{-r/\beta}$. The typical sample paths of $X(t)$ and $Y(t)$ for different values of k_{on} and k_{off} are shown in Figs. 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.

bioRxiv preprint doi: https://doi.org/10.1101/316166; this version posted May 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Analysis

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

$$\mathrm{d}Y_-(t) = -\gamma Y_-(t)\mathrm{d}t + \mathrm{dC}_-(t;\lambda_-,\beta), \tag{5}$$
$$\mathrm{d}Y_+(t) = -\gamma Y_+(t)\mathrm{d}t + X(t)\mathrm{dC}_+(t;\lambda_+,\beta), \tag{6}$$

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 procedure in references (41, 42), we obtain the following system of integro-differential equations along with the normalization condition:

$$\frac{\partial p_-(\tau,y)}{\partial \tau} = \frac{\partial}{\partial y}(y p_-(\tau,y)) + \Lambda_-\left(\int_0^y \overline{q}(y')p_-(\tau,y-y')\mathrm{d}y' - p_-(\tau,y)\right), \tag{7}$$
$$\frac{\partial p_+(\tau,0,y)}{\partial \tau} = \frac{\partial}{\partial y}(y p_+(\tau,0,y)) - K_{\mathrm{on}}p_+(\tau,0,y) + K_{\mathrm{off}}p_+(\tau,1,y), \tag{8}$$
$$\frac{\partial p_+(\tau,1,y)}{\partial \tau} = \frac{\partial}{\partial y}(y p_+(\tau,1,y)) + K_{\mathrm{on}}p_+(\tau,0,y) - K_{\mathrm{off}}p_+(\tau,1,y) + \Lambda_+\left(\int_0^y \overline{q}(y')p_+(\tau,1,y-y')\mathrm{d}y' - p_+(\tau,1,y)\right), \tag{9}$$
$$\int_0^\infty p_+(\tau,y)\mathrm{d}y \equiv 1, \tag{10}$$

where $\tau = \gamma t$, $K_{\mathrm{on}} = k_{\mathrm{on}}/\gamma$, $K_{\mathrm{off}} = k_{\mathrm{off}}/\gamma$, and $\Lambda_i = \lambda_i/\gamma$ ($i \in \{+,-\}$). Note that Eq. (7) is the so-called Kolmogorov-Feller equation (44) introduced by Friedman and his coworkers as a model of constitutive gene expression (26).

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

$$\overline{p}_-(y) = \frac{y^{\Lambda_- -1}\mathrm{e}^{-y/\beta}}{\Gamma(\Lambda_-)\beta^{\Lambda_-}}, \tag{11}$$
$$\overline{p}_+(x,y) = x + (1-2x)C_0\int_0^1 \frac{y^{a-1}\mathrm{e}^{-y/(\beta w)}}{\Gamma(a)(\beta w)^a}\cdot\frac{w^{b-1}(1-w)^{c-b}}{B(b,c-b+1)}\mathrm{d}w, \tag{12}$$
$$\overline{p}_+(y) = \int_0^1 \frac{y^{a-1}\mathrm{e}^{-y/(\beta w)}}{\Gamma(a)(\beta w)^a}\cdot\frac{w^{b-1}(1-w)^{c-b-1}}{B(b,c-b)}\mathrm{d}w, \tag{13}$$
$$\overline{p}(y) = \int_0^y \overline{p}_-(y')\overline{p}_+(y-y')\mathrm{d}y', \tag{14}$$

where $C_0 = K_{\mathrm{off}}/(K_{\mathrm{on}} + K_{\mathrm{off}})$, $a + b = K_{\mathrm{on}} + K_{\mathrm{off}} + \Lambda_+$, $ab = K_{\mathrm{on}}\Lambda_+$, and $c = K_{\mathrm{on}} + K_{\mathrm{off}}$ (see the Supplemental Material 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 $K_{\mathrm{on}}, K_{\mathrm{off}} \ll 1$, Eq. (13) can be approximated as follows:

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

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:

$$\overline{p}_i(ky;\boldsymbol{\theta}_i)(k\mathrm{d}y) = \overline{p}_i(y;\tilde{\boldsymbol{\theta}}_i)\mathrm{d}y, \tag{16}$$

where $\boldsymbol{\theta}_i = (k_{\mathrm{on}}, k_{\mathrm{off}}, \lambda_i, \beta, \gamma)$ and $\tilde{\boldsymbol{\theta}}_i = (k_{\mathrm{on}}, k_{\mathrm{off}}, \lambda_i, \beta/k, \gamma)$ ($i \in \{+,-\}$). This result indicates that Eqs. (11)–(14) can be fit to various experimental data with an arbitrary measurement scale of y. 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 (45, 46). Through the analysis, we can also calculate any order of moment (see the Supplemental Material for the calculation).

bioRxiv preprint doi: https://doi.org/10.1101/316166; this version posted May 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

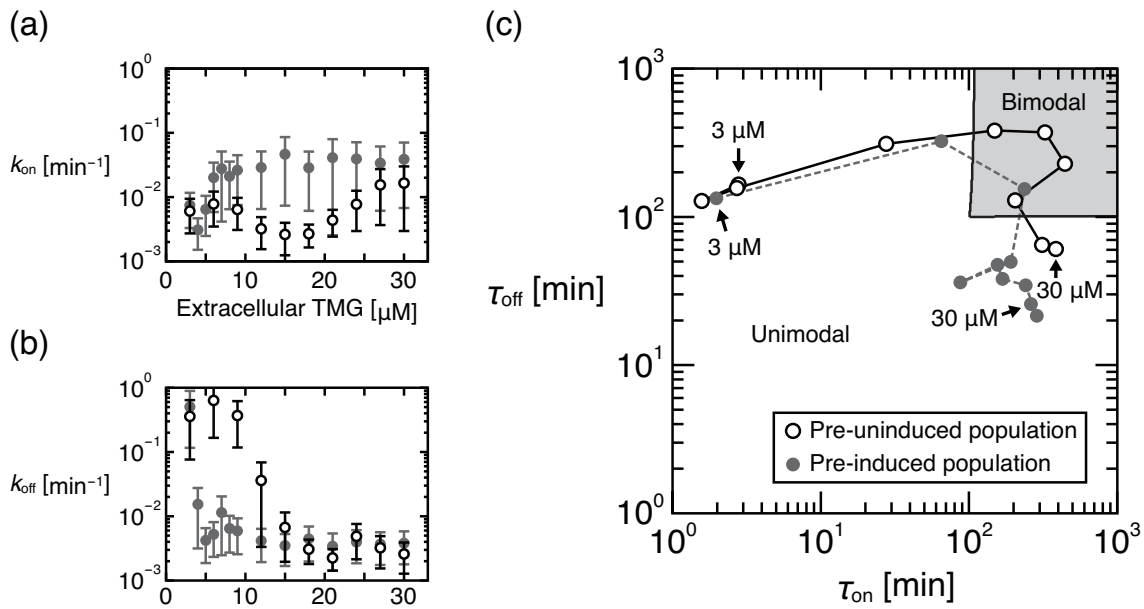

Figure 2: The estimated mean values of our model parameters k_{on} and k_{off} for the preuninduced (unfilled circles) and preinduced (filled circles) populations toward the various extracellular TMG concentrations. (a) and (b) The rate parameters k_{on} and k_{off} versus TMG levels. (c) The average 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 $\overline{p}(y)$. The solid and dashed lines connecting the points are simply visual guides.

RESULTS

Finally, we apply our theory to the TMG-induced system of *lacZYA* expression in *E. coli*, as reported in 2004 (15). 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) in each cell toward the various TMG concentrations was measured. The remaining definitions and assumptions in this study are as follows: (i) x, y, and z denote the activity of the *lac* promoter and the intracellular levels of the GFP and TMG, respectively; (ii) z is constant for the system, which is assumed to be in a steady state (15); (iii) $\overline{p}(y)$ denotes the cell-to-cell variation of y at the steady state; (iv) $k_{on} \gg k_{off}$ is defined with respect to the fully induced population; (v) $\lambda_- = 0$ by assuming that the genetic switch is all-or-none; (vi) $\lambda_- + \lambda_+$, β, and γ are the maximum transcription rate constant 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; and (vii) (k_{on}, k_{off}) and (λ_+, β) are dependent on and independent of the extracellular TMG, respectively. Based on assumptions (i)–(vii), we estimate λ_+ and β from the experimental data with the fully induced population (15, 43, 47–49). Here, the estimated values are $\lambda_+ = 15$ min^{-1}, $\beta = 0.074$, and $\gamma = 0.01$ min^{-1} (see the Supplemental Material for the parameter estimation).

As shown in the previous study, most of the GFP distributions are far from Gaussian, which indicates that such distributions are poorly characterized by the mean and variance. Hence, we adopt a Bayesian approach utilizing the Metropolis algorithm (50), and we estimate the posterior distributions of k_{on} and k_{off} from the published single-cell data (15). Note that each posterior had a single peak over our searched parameter range. Figs. 2(a) and (b) show the estimated mean values of k_{on} and k_{off} as functions of the extracellular TMG levels, and Fig. 2(c) shows the values of τ_{on} ($=k_{off}^{-1}$) and τ_{off} ($=k_{on}^{-1}$), 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_{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. 2(c)). This result is comparable with those found at the mRNA level (47). In addition, we found that the preuninduced and preinduced cells mainly modulate k_{on} for 24–30 and 6–30 μM TMG, respectively, the values of which are necessary for their further induction.

bioRxiv preprint doi: https://doi.org/10.1101/316166; this version posted May 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

DISCUSSION

We have shown that our theoretical framework can estimate the biological parameters from single-cell data measured as fluorescence intensity. Fig. 2(c) shows that $\tau_{\mathrm{on}} \approx 2.78$ and $\tau_{\mathrm{off}} \approx 164.82$ 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 gene expression burst lasts ∼3 to 15 min and that the average time between two adjacent expressions is 46 min (43). Hence, careful tuning of the parameters λ_i ($i \in \{+, -\}$), β, and γ in accordance with given experimental data may help researchers accurately quantify the kinetic parameters k_{on} and k_{off}.

Interestingly, Figs. 2(a) and (b) show that $k_{\mathrm{on}}(z)$ and $k_{\mathrm{off}}(z)$ are nonlinear functions of TMG levels. In this case, if z 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 derivation provided that $k_{\mathrm{on}}(z)$ and $k_{\mathrm{off}}(z)$ are convex functions. However, Figs. 2(a) and (b) indicate that k_{on} and k_{off} are nonconvex, which leads to a notorious mixed convex problem (51).

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 (52). Gnugge et al. reported that *E. coli* and *S. cerevisiae* have a similar core network in the lactose and galactose utilizing systems, respectively (53), 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 (Fig .2(c)), which may provide a better understanding of the quantitative differences across species. Dar, Razooky et al. proposed an ingenious method for estimating their model parameters from three statistics (i.e., the coefficient of variation, expression level, and autocorrelation time) and mapped the kinetic features of human gene expression into their parameter space through a genome-wide experiment (54). However, their method is based on a finite number of statistics, which can estimate a smaller number of parameters, and the experiment requires a high-resolution real-time monitoring, which is costly. To overcome these issues, a Bayesian or maximum likelihood approach is suitable, as Shahrezaei et al. mentioned in their report (27). In this respect, our framework including MCMC performed well without any concern for the number of statistics. Choi, Cai et al. experimentally examined various *E. coli* strains with different genetic constructions and found that the binding affinity of the effector protein to its target loci significantly changes the protein distribution (17). Linking DNA structure, such as looping, chemical modification, and DNA-protein complex formation, to its expression pattern is an important future work.

CONCLUSION

We have proposed a fairly general model of a gene regulatory system along with the clear biophysical assumptions, and we formulated it by both stochastic differential equations (Eqs. (2) and (3)) and the corresponding master equations (Eqs. (7)–(10)). We subsequently derived the 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.

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. 2(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 (1, 55), cell differentiations (13), and evolution (22).

AUTHOR CONTRIBUTIONS

K.I. performed mathematical modeling, analysis, and parameter estimation. N.O., Y.K., and K.I. developed the strategy. Y.K. initiated the project. K.I. and N.O. wrote the article.

ACKNOWLEDGMENTS

The authors thank Satoshi Mizuno, Soichi Ogishima, Naoko Kasahara, Satoshi Nagaie, and Rie Norita for their helpful discussion. This work was partially supported by JSPS KAKENHI Grant Number 16H03939.

bioRxiv preprint doi: https://doi.org/10.1101/316166; this version posted May 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

REFERENCES

1. Elowitz, M. B., and S. Leibler, 2000. A synthetic oscillatory network of transcriptional regulators. *Nature* 403:335–338.
2. Golding, I., J. Paulsson, S. M. Zawilski, and E. C. Cox, 2005. Real-time kinetics of gene activity in individual bacteria. *Cell* 123:1025–1036.
3. Chubb, J. R., T. Trcek, S. M. Shenoy, and R. H. Singer, 2006. Transcriptional pulsing of a developmental gene. *Curr. Biol.* 16:1018–1025.
4. Gardner, T. S., C. R. Cantor, and J. J. Collins, 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403:339–342.
5. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain, 2002. Stochastic gene expression in a single cell. *Science* 297:1183–1186.
6. Blake, W. J., M. Kærn, C. R. Cantor, and J. J. Collins, 2003. Noise in eukaryotic gene expression. *Nature* 422:633–637.
7. Zenklusen, D., D. R. Larson, and R. H. Singer, 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. *Nat. Struct. Mol. Biol.* 15:1263–1271.
8. 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.
9. Zamparo, M., F. Chianale, C. Tebaldi, M. Cosentino-Lagomarsino, M. Nicodemi, and A. Gamba, 2015. Dynamic membrane patterning, signal localization and polarity in living cells. *Soft Matter* 11:838–849.
10. Yamanaka, Y., F. Lanner, and J. Rossant, 2010. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development* 137:715–724.
11. Acar, D., A. Becskei, and A. van Oudenaarden, 2005. Enhancement of cellular memory by reducing stochastic transitions. *Nature* 435:228–232.
12. Becskei, A., B. Séraphin, and L. Serrano, 2001. Positive feedback in eukaryotic gene networks: cell differentiation by graded to binary response conversion. *EMBO J.* 20:2528–2535.
13. Süel, G. M., J. Garcia-Ojalvo, L. M. Liberman, and M. B. Elowitz, 2006. An excitable gene regulatory circuit induces transient cellular differentiation. *Nature* 440:545–550.
14. Fraser, D., and M. Kærn, 2009. A chance at survival: gene expression noise and phenotypic diversification strategies. *Mol. Microbiol.* 71:1333–1340.
15. Ozbudak, E. M., M. Thattai, H. N. Lim, B. I. Shraiman, and A. van Oudenaarden, 2004. Multistability in the lactose utilization network of *Escherichia coli*. *Nature* 427:737–740.
16. Mettetal, J. T., D. Muzzey, J. M. Pedraza, E. M. Ozbudak, and A. van Oudenaarden, 2006. Predicting stochastic gene expression dynamics in single cells. *Proc. Natl. Acad. Sci. USA.* 103:7304–7309.
17. Choi, P. J., L. Cai, K. Frieda, and X. S. Xie, 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science* 322:442–446.
18. Acar, M., B. F. Pando, F. H. Arnold, M. B. Elowitz, and A. van Oudenaarden, 2010. A General mechanism for network-dosage compensation in gene circuits. *Science* 329:1656–1660.
19. Peng, W., P. Liu, Y. Xue, and M. Acar, 2015. Evolution of gene network activity by tuning the strength of negative-feedback regulation. *Nat. Commun.* 6:1–9.
20. Peng, W., R. Song, and M. Acar, 2016. Noise reduction facilitated by dosage compensation in gene networks. *Nat. Commun.* 7:1–8.
21. Blake, W. J., G. Balázsi, M. A. Kohanski, F. J. Isaacs, K. F. Murphy, Y. Kuang, C. R. Cantor, D. R. Walt, and J. J. Collins, 2006. Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* 24:853–865.

bioRxiv preprint doi: https://doi.org/10.1101/316166; this version posted May 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

23. Peccoud, J., and B. Ycart, 1995. Markovian modelling of gene product synthesis. *Theor. Popul. Biol.* 48:222–234.

24. Paulsson, J., and M. Ehrenberg, 2000. Random signal fluctuations can reduce random fluctuations in regulated components of chemical regulatory networks. *Phys. Rev. Lett.* 84:5447–5450.

25. Hornos, J. E. M., D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, 2005. Self-regulating gene: An exact solution. *Phys. Rev. E* 72:051907.

26. Friedman, N., L. Cai, and X. S. Xie, 2006. Linking stochastic dynamics to population distribution: an analytical framework of gene expression. *Phys. Rev. Lett.* 97:168302.

27. Shahrezaei, V., and P. S. Swain, 2008. Analytical distributions for stochastic gene express. *Proc. Natl. Acad. Sci. USA.* 105:17256–17261.

28. Assaf, M., E. Roberts, and Z. L. Schulten, 2011. Determining the stability of genetic switches: explicitly accounting for mrna noise. *Phys. Rev. Lett.* 106:248102.

29. Schwabe, A., K. N. Rybakova, and F. J. Bruggeman, 2012. Transcription stochasticity of complex gene regulation models. *Biophys. J.* 103:1152–1161.

30. Bokes, P., J. R. King, A. T. A. Wood, and M. Loose, 2012. Exact and approximate distributions of protein and mRNA levels in the low-copy regime of gene expression. *J. Math. Biol.* 64:829–854.

31. Vandecan, Y., and R. Blossey, 2013. Self-regulatory gene: An exact solution for the gene gate model. *Phys. Rev. E* 87:042705.

32. Pendar, H., T. Platini, and R. V. Kulkarni, 2013. Exact protein distributions for stochastic models of gene expression using partitioning of Poisson processes. *Phys. Rev. E* 87:042720.

33. Zhang, J., and T. Zhou, 2014. Promoter-mediated transcriptional dynamics. *Biophys. J.* 106:479–488.

34. Zhou, T., and T. Liu, 2015. Quantitative analysis of gene expression systems. *Quant. Biol.* 3:168–181.

35. Suter, D. M., N. Molina, D. Gatfield, K. Schneider, U. Schibler, and F. Naef, 2011. Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332:472–474.

36. Bakstad, D., A. Adamson, D. G. Spiller, and M. R. H. White, 2012. Quantitative measurement of single cell dynamics. *Curr. Opin. Biotech.* 23:103–109.

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

38. Wallace, E. W. J., D. T. Gillespie, K. R. Sanft, and L. R. Petzold, 2012. Linear noise approximation is valid over limited times for any chemical system that is sufficiently large. *IET Syst. Biol.* 6:102–115.

39. Yagil, G., and E. Yagil, 1971. On the relation between effector concentration and the rate of induced enzyme synthesis. *Biophys. J.* 11:11–27.

40. Popović, N., C. Marr, and P. S. Swain, 2016. A geometric analysis of fast-slow models for stochastic gene expression. *J. Math. Biol.* 72:87–122.

41. 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.

42. Denisov, S. I., W. Horsthemke, and P. Hänggi, 2009. Generalized Fokker-Planck equation: derivation and exact solutions. *Eur. Phys. J. B* 68:567–575.

43. Yu, J., J. Xiao, X. Ren, K. Lao, and X. S. Xie, 2006. Probing gene expression in live cells, one protein molecule at a time. *Science* 311:1600–1603.

bioRxiv preprint doi: https://doi.org/10.1101/316166; this version posted May 7, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

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

45. Kim, J. K., and J. C. Marioni, 2013. Inferring the kinetics of stochastic gene expression from single-cell RNA-sequencing data. *Genome Biol.* 14:1–12.

46. Iida, K., and Y. Kimura, 2015. Springer, New York, volume 1, chapter Mathematical theory to compute stochastic cellular processes, 117–120. 1 edition.

47. So, L., A. Ghosh, C. Zong, L. A. Sepúlveda, R. Segev, and I. Golding, 2011. General properties of transcriptional time series in *Escherichia coli*. *Nat. Genet.* 43:554–560.

48. Leive, L., and V. Kollin, 1967. Synthesis, utilization and degradation of lactose operon mRNA in *Escherichia coli*. *J. Mol. Biol.* 24:247–269.

49. Kennell, D., and H. Riezman, 1977. Transcription and translation initiation frequencies of the *Escherichia coli lac* operon. *J. Mol. Biol.* 114:1–21.

50. Brooks, S. P., 1998. Markov chain Monte Carlo method and its application. *J. Roy. Stat. Soc. D-Sta.* 47:69–100.

51. Niculescu, C. P., and I. Rovenţa, 2015. Relative convexity and its applications. *Aequat. Math.* 89:1389–1400.

52. Lionnet, T., and R. H. Singer, 2012. Transcription goes digital. *EMBO Rep.* 13:313–321.

53. Gnügge, R., L. Dharmarajan, M. Lang, and J. Stelling, 2016. An orthogonal permease-inducer-repressor feedback loop shows bistability. *ACS Synth. Biol.* 5:1098–1107.

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

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

SUPPLEMENTARY MATERIAL

An online supplement to this article can be found by visiting BJ Online at http://www.biophysj.org.