

Approximation scheme based on effective interactions for stochastic gene regulation

Jun Ohkubo*

Graduate School of Informatics, Kyoto University, 36-1, Yoshida Hon-machi, Sakyo-ku, Kyoto-shi, Kyoto 606-8501, Japan

(Received 13 December 2010; revised manuscript received 16 February 2011; published 20 April 2011)

Since gene regulatory systems sometimes contain only a small number of molecules, these systems are not described well by macroscopic rate equations; a master equation approach is needed for such cases. We develop an approximation scheme for dealing with the stochasticity of the gene regulatory systems. Using an effective interaction concept, original master equations can be reduced to simpler master equations, which can be solved analytically. We apply the approximation scheme to self-regulating systems with monomer or dimer interactions, and a two-gene system with an exclusive switch. The approximation scheme can recover the bistability of the exclusive switch adequately.

DOI: [10.1103/PhysRevE.83.041915](https://doi.org/10.1103/PhysRevE.83.041915)

PACS number(s): 87.10.Ca, 02.50.Ey, 82.39.-k

I. INTRODUCTION

Recently, the stochastic nature of small systems has attracted much attention [1–3]. One of the interesting examples of the stochasticity is a gene regulatory system; it has been known experimentally that the gene regulatory systems show various phenomena caused by intrinsic noise [4,5]. The gene regulatory systems basically consist of genes, RNA, and proteins. The genes could sometimes be activated or repressed by regulatory proteins known as transcription factors. The number of regulatory proteins is sometimes very small, and there are large fluctuations. From a theoretical point of view, the gene regulatory systems have been studied a lot using Monte Carlo simulations (e.g., Refs. [6,7]). In addition, in order to gain insights into mechanisms or functions of the gene regulatory systems, many analytical studies have been done [8–18]. For example, for a self-regulating system with monomer binding interactions, an exact solution is already known [12]. However, when one considers more complicated systems, some approximations are needed. Such approximations have also been developed: a Fokker-Planck or Langevin equation approach [8–10], a variational approach [11,14,15], and a self-consistent proteomic field approximation [13].

A gene regulatory system with only two genes and feedback mechanisms has been studied a lot because it plays an important role as a genetic switch; two distinct stable states emerge, and they could be switched either spontaneously or by external signals. In a mathematical description for the gene regulatory systems, the RNA is sometimes neglected for simplicity, and only genes and regulatory proteins are considered. When we construct a macroscopic rate equation, in which fluctuations in protein copy numbers or gene expression states are neglected, the analysis for the rate equation tells us the following facts: A system with two mutually repressing genes shows a bistability, and cooperative binding of regulatory proteins is important for making the bistability [19,20]. Here the cooperative binding means that combinations of two or more proteins are needed to activate or repress genes. The macroscopic rate equation gives multiple stable solutions,

and each solution corresponds to a stable state of the gene regulatory systems, which causes the bistability. Hence, for the cooperative binding cases, it may be enough to use the macroscopic treatments in order to investigate the qualitative behavior of the bistability. However, other studies have shown that a so-called exclusive switch shows a bistability even when the macroscopic rate equations have only one solution [21–23]. Although the bistability has been confirmed numerically using Monte Carlo simulations, no exact or approximated analytical treatment has been proposed yet, to the best of our knowledge.

In the present paper, we develop a new approximation scheme for gene regulatory systems. In the approximation scheme, there is no need to use a continuous description such as Fokker-Planck or Langevin equations, and hence the smallness or discrete properties of the system are not neglected. The basic idea of the approximation is similar to the “self-consistent proteomic field approximation” developed by Walczak *et al.* [13]. In the self-consistent proteomic field approximation, a joint probability for all genes is approximated as a product of probability distributions for each gene, and then the interactions between genes and regulatory proteins can be evaluated “exactly” within this approximation. While the self-consistent proteomic field approximation was applied to toggle switches, further approximations are needed for a kind of regulating system, such as self-regulating systems, as discussed in Ref. [13]. We here develop a more applicable approximation scheme; the interactions between a gene and regulatory proteins are approximated first, and effective interactions are introduced. The new approximation scheme enables us to give analytical expressions for probability distributions of the numbers of proteins without loss of the discreteness property of the system. We will show that the new approximation scheme is applicable not only to the self-regulating systems, but also to the exclusive switch without cooperative interactions.

The present paper is constructed as follows. In Sec. II, we give a brief review of a stochastic model for gene regulation. In Sec. III, self-regulating systems are studied. Section III B gives one of the important results in the present paper, in which our approximation scheme is proposed. The proposed approximation scheme is applied to the exclusive switch in Sec. VI. Section V gives concluding remarks.

*ohkubo@i.kyoto-u.ac.jp

II. STOCHASTIC MODEL FOR GENE REGULATION

We here briefly review the basic biology of gene regulatory system and a simplified stochastic model, for readers' convenience.

A gene regulatory system consists of many components, such as genes, RNA, and proteins. The transcription of a gene is initiated by a binding of RNA polymerase to a promoter site of the gene in the DNA. The binding of regulatory proteins (or molecules), so-called transcription factors, regulates the transcription initiation. These regulatory proteins bind to their own target operator sites, and they sometimes act as repressors (which repress the transcription) or activators (which enhance the transcription). When the RNA polymerase binds to a gene, the gene sequence is transferred to a messenger RNA (mRNA), and the mRNA is translated into a protein molecule by a ribosome enzyme complex. The produced proteins play various roles, and sometimes they can become regulatory signals for genes.

Although all of the above reactions would be important for the gene regulatory systems, mRNA is sometimes neglected in stochastic modeling for simplicity. That is, the translation from mRNA to proteins is straightforward, and then we assume that an activated gene directly increases the number of proteins. In addition, we consider that a repressed gene cannot produce any proteins, which makes analytical treatments much simpler [12].

In the present paper, all regulatory proteins act as repressors. If regulatory proteins are not binding to a gene, we call a state of the gene an "ON" state; if not, the gene is in an "OFF" state. A gene in an OFF state cannot produce any proteins, as we assumed above.

III. SELF-REGULATING SYSTEM

In this section, first, we propose and explain a new approximation scheme using a simplest model, i.e., a self-regulating system with monomer interactions. Results obtained from the approximation scheme will be compared with exact solutions studied previously [6,12]. Second, the scheme will be applied to a self-regulating system with dimer interactions, and comparisons with Monte Carlo simulations will be given.

A. Model

First, we give a brief explanation for a self-regulating system. In the self-regulating system, there is only one gene, and it produces proteins. The produced proteins are considered as regulatory proteins for the gene, and the regulation is a repressed one. In this sense, there is a self-regulation mechanism. Figure 1 shows the self-regulating system. When the gene is in ON state, it produces the regulatory proteins with rate g . The degradation rate of the regulatory proteins is k . The regulatory proteins can bind the gene with a rate function $H(n)$, where n is the number of "free" regulatory proteins. The function $H(n)$ can be a complicated function of the regulatory proteins; e.g., $H(n) = hn$ for monomer interactions, and $H(n) = hn(n - 1)/2$ for dimer interactions, where h is a rate constant for the binding. f is a rate constant with which the regulatory protein is released from the repressor site of the gene.

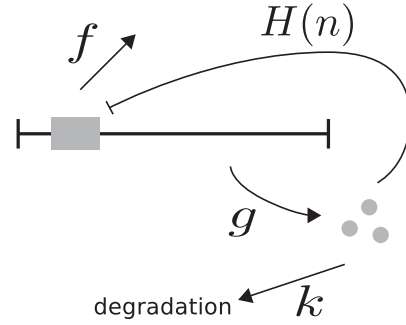


FIG. 1. A schematic illustration of the self-regulating gene.

B. Approximation scheme

We here consider the simplest interaction case, i.e., a monomer interaction case. Hence, $\mathcal{H}(n)$ in Fig. 1 is written as hn , as discussed in Sec. III A. For the monomer interaction cases, exact solutions have already been known [6,12], and therefore we can compare our approximation results with the exact solutions for the monomer interaction case.

We first note one assumption in order to make analytical treatments simpler. The assumption has also been used in a previous work [6]; i.e., one of the proteins in an ON state is inert. The inert protein cannot repress the gene, and it is not degraded. Although there are a few differences between usual stochastic simulations and this analytical treatment, it has already been discussed that the assumption gives quantitatively good results [6]. Hence, we here employ this assumption. Figure 2 shows the transition scheme for the usual stochastic simulations and the analytical treatment. α_n and β_n correspond to states in which there are n regulatory proteins for ON and OFF states, respectively. In the usual stochastic simulations, the degradation rate of the proteins, i.e., the change from α_n to α_{n-1} , is proportional to the number of proteins n . In contrast,

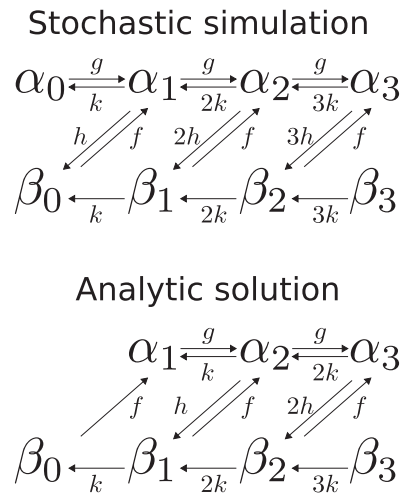


FIG. 2. Transition scheme for usual simulations and analytical calculations for monomer interaction cases. α_n and β_n correspond to states with which there are n regulatory proteins for ON and OFF states, respectively. In the analytical calculations, one of the proteins is considered as an inert one when the gene is in an ON state, and the inert protein is also included in α_n ; the number of "free" regulatory proteins in an ON state is $n - 1$.

the above assumption means that the rate from α_n to α_{n-1} in the analytical treatment is proportional to $n - 1$, not to n . For convenience, we introduce a new notation α'_n , where there are n “free” regulatory proteins for an ON state; $\alpha_{n+1} \equiv \alpha'_n$. Note that the inert protein is not a “free” regulatory protein, and α'_n does not include the inert protein.

Here we propose a new approximation scheme. The key of the approximation scheme is to use “an effective interaction.” The effective interaction means that the interaction factor in Fig. 1, $H(n)$, is replaced as a scalar value; i.e., $\mathcal{H}(n) = \tilde{h}$. Note that the effective interaction \tilde{h} is not a function of the regulatory proteins, but a constant. Hence, the master equation for this approximated system is written as follows:

$$\begin{aligned} \frac{dP(\alpha'_n, t)}{dt} = & g[P(\alpha'_{n-1}, t) - P(\alpha'_n, t)] \\ & + k[(n+1)P(\alpha'_{n+1}, t) - nP(\alpha'_n, t)] \\ & - \tilde{h}P(\alpha'_n, t) + fP(\beta_n, t), \end{aligned} \quad (1)$$

$$\begin{aligned} \frac{dP(\beta_n, t)}{dt} = & k[(n+1)P(\beta_{n+1}, t) - nP(\beta_n, t)] \\ & + \tilde{h}P(\alpha'_n, t) - fP(\beta_n, t), \end{aligned} \quad (2)$$

where $P(\alpha'_n, t)$ and $P(\beta_n, t)$ are probabilities for n free regulatory proteins for ON and OFF states, respectively. Note that $P(\alpha_{n+1}, t) \equiv P(\alpha'_n, t)$.

Because the interaction factor $H(n) = \tilde{h}$ has a simple form, the analytic solution can be easily calculated by using the generating function approach, as in the case of the exact solutions [6, 12]. The generating functions for stationary states are defined as follows:

$$\alpha'(z) = \sum_{n=0}^{\infty} P(\alpha'_n) z^n, \quad (3)$$

$$\beta(z) = \sum_{n=0}^{\infty} P(\beta_n) z^n, \quad (4)$$

where $P(\alpha'_n) = \lim_{t \rightarrow \infty} P(\alpha'_n, t)$ and $P(\beta_n) = \lim_{t \rightarrow \infty} P(\beta_n, t)$. We note that $\alpha'(1) + \beta(1) = 1$ from the normalization condition. Using the generating functions, various information about the number of proteins can be obtained. For example, the probability $P(\alpha'_n)$ is calculated from

$$P(\alpha'_n) = \frac{1}{n!} \left. \frac{\partial^n \alpha'(z)}{\partial z^n} \right|_{z=0}. \quad (5)$$

It is also possible to calculate moments or cumulants. Since the probability with which the gene is in an ON state is given as $\alpha'(1)$. The number of total proteins in the system is calculated as

$$\langle n \rangle = \left. \frac{\partial \alpha'(z)}{\partial z} \right|_{z=1} + 1 \times \alpha'(1) + \left. \frac{\partial \beta(z)}{\partial z} \right|_{z=1}, \quad (6)$$

where the second term in the right-hand side means a contribution from the inert protein in an ON state.

Putting the left-hand sides of Eqs. (1) and (2) as zero and rewriting Eqs. (1) and (2) in terms of the generating functions $\alpha'(z)$ and $\beta(z)$, stationary solutions for the generating functions

are obtained as follows:

$$\alpha'(z) = A F[a, b, N(z-1)], \quad (7)$$

$$\beta(z) = \left(1 + \frac{\tilde{h}}{f}\right) A F[a-1, b-1, N(z-1)] - \alpha'(z), \quad (8)$$

where $A = f/(f + \tilde{h})$ and

$$N = \frac{g}{k}, \quad a = 1 + \frac{f}{k}, \quad b = 1 + \frac{f + \tilde{h}}{k}. \quad (9)$$

$F(p, q, r)$ is the Kummer confluent hypergeometric function,

$$F(p, q, r) \equiv \sum_{n=0}^{\infty} \frac{(p)_n r^n}{(q)_n n!}, \quad (10)$$

where $(p)_n = p(p+1)(p+2) \cdots (p+n-1)$.

A remaining task is to determine the effective interaction \tilde{h} . For the self-regulating system with monomer interactions, the binding of the regulatory proteins occurs only when the system is in an ON state. Hence, the number of proteins, which can be attached to the binding site, should be equal to the number of free proteins for an ON state.

According to the above discussions, we here set the effective interaction \tilde{h} as

$$\tilde{h} = h \langle n \rangle_{\alpha'}, \quad (11)$$

where $\langle n \rangle_{\alpha'}$ is the expectation of the number of free regulatory proteins under a condition that the gene is in an ON state (conditional expectation). Because it is possible to evaluate the conditional expectation using the generating function as

$$\langle n \rangle_{\alpha'} \equiv \frac{1}{\alpha'(1)} \left. \frac{\partial \alpha'(z)}{\partial z} \right|_{z=1} = \frac{g(k+f)}{k(k+f+\tilde{h})}, \quad (12)$$

we obtain the following self-consistent equation by inserting Eq. (12) into Eq. (11):

$$\tilde{h} = h \frac{g(k+f)}{k(k+f+\tilde{h})}. \quad (13)$$

Solving Eq. (13), we finally obtain

$$\tilde{h} = \frac{-(k^2 + kf) + \sqrt{(k^2 + kf)^2 + 3khg(k+f)}}{2k}. \quad (14)$$

Once the effective interaction \tilde{h} is determined, all statistical properties related to the number of regulatory proteins are evaluated from the generating functions, as denoted above. For example, using the following property of the Kummer confluent hypergeometric function,

$$\frac{\partial}{\partial r} F(p, q, r) = \frac{p}{q} F(p+1, q+1, r), \quad (15)$$

the probability distributions for the numbers of free proteins are written as

$$P(\alpha'_n) = \frac{A}{n!} N^n \frac{(a)_n}{(b)_n} F(a+n, b+n, -N), \quad (16)$$

$$\begin{aligned} P(\beta_n) = & \frac{A}{n!} \left(1 + \frac{\tilde{h}}{f}\right) N^n \frac{(a-1)_n}{(b-1)_n} \\ & \times F(a-1+n, b-1+n, -N) - P(\alpha'_n). \end{aligned} \quad (17)$$

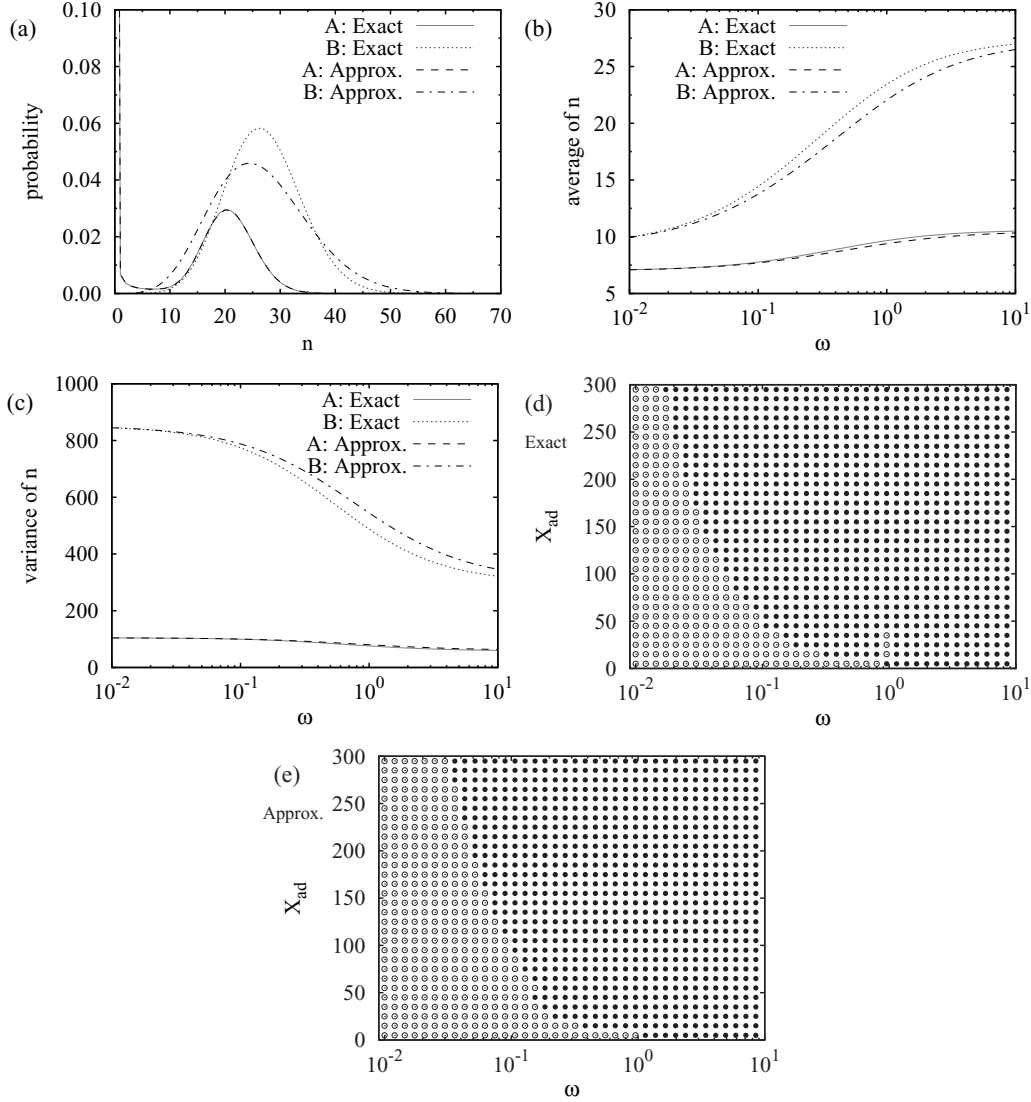


FIG. 3. Comparison between results from the exact solutions and those from the approximation scheme. (a) Probability distributions of the number of proteins. $X_{eq} = 10.0$, $X_{ad} = 10.0$, $\omega = 0.01$ in Case A; $X_{eq} = 10.0$, $X_{ad} = 50.0$, $\omega = 10.0$ in Case B. (b) and (c) Averages and variances of the number of proteins, respectively. Only the value of ω was changed, and $X_{eq} = 10.0$, $X_{ad} = 10.0$ in Case A; $X_{eq} = 10.0$, $X_{ad} = 50.0$ in Case B. (d) and (e) Results for modality in the exact solutions and in the approximation scheme, respectively. The filled circle means unimodal regions, and the empty circles indicate regions with bimodal distributions. $X_{eq} = 10.0$ for (d) and (e).

C. Results for monomer interactions

For monomer interaction cases, the exact solutions are obtained [6,12]. Hence, we here compare the exact results with results obtained by the approximation scheme.

For the comparison, we here introduce rescaled parameters as follows [6,12]:

$$\omega = \frac{f}{k}, \quad X_{eq} = \frac{f}{h}, \quad X_{ad} = \frac{g}{2k}, \quad (18)$$

and, for simplicity, we set $k = 1$ in all numerical evaluations. These rescaled parameters are helpful to understand properties of the genetic switch. The parameter X_{ad} characterizes the synthesis and degradation processes, and large X_{ad} would give a large average number of proteins. X_{eq} is related to the equilibrium constant of the binding and unbinding process. Finally, ω is a parameter called a ‘‘adiabaticity parameter.’’ ω measures how rapidly the gene can equilibrate in a gene state.

If ω is small, the synthesis and degradation behaves almost like an independent birth and death process, and there would be two peaks corresponding to the binding and unbinding states, respectively. For details of these parameters, for example, see Ref. [6].

First, the probability distributions of the number of protein were compared. Figure 3(a) shows the results. Here we performed two cases: In Case A, we set $X_{eq} = 10.0$, $X_{ad} = 10.0$, and $\omega = 0.01$; in Case B, $X_{eq} = 10.0$, $X_{ad} = 50.0$, and $\omega = 10.0$. For Case A, the results obtained from the approximation scheme give good agreement with those from the exact solutions, and it is difficult to see the difference. Although there are quantitative differences between the exact and approximate solutions for Case B, the approximation scheme gives qualitatively good results despite the rough approximation. Figures 3(b) and 3(c) show the averages and variances of the number of proteins for various values of ω

when $X_{\text{eq}} = 10.0$ and $X_{\text{ad}} = 10.0$ for Case A and $X_{\text{eq}} = 10.0$ and $X_{\text{ad}} = 50.0$ for Case B, respectively. Figures 3(d) and 3(e) show results for modality in the exact solutions and in the approximation scheme, respectively, for $X_{\text{eq}} = 10.0$. These results indicate that the approximation scheme works well, at least qualitatively.

D. Results for dimer interactions

As a second example, a self-regulating system with dimer interactions is studied. The dimer interaction, i.e., $H(n) = hn(n-1)/2$, has a more complicated form than the monomer interaction cases, and exact solutions for the dimer interaction cases have not been known yet. There are some numerical studies for the dimer interaction cases (for example, see Ref. [24]), and here we compare results obtained from the approximation scheme with those of Monte Carlo simulations using the Gillespie algorithm [25].

In the dimer interaction cases, the transition scheme for analytical calculations is different from the monomer interaction cases; see Fig. 4. The master equations are the same as Eqs. (1) and (2), but the effective interaction should be set as

$$\tilde{h} = h \frac{(n(n-1))_{\alpha'}}{2}, \quad (19)$$

and we should interpret α'_n as $P(\alpha_{n+2}, t) = P(\alpha'_n, t)$.

Using the similar procedure written in Sec. III B, the effective interaction \tilde{h} is obtained by solving the following self-consistent equation:

$$\tilde{h} = \frac{h}{2} \frac{1}{\alpha'(1)} \left. \frac{\partial^2}{\partial z^2} \alpha'(z) \right|_{z=0}. \quad (20)$$

Since it is a slightly complicated task to obtain the analytical expression for the effective interaction \tilde{h} , we numerically solved the self-consistent equation (20). Once we evaluate the effective interaction \tilde{h} , all properties related to the number of proteins are immediately obtained, in a similar way to Sec. III B.

As in the case of Fig. 3, we evaluated the probability distributions, the averages and variances of the number of proteins, and the modality. In Case A in Figs. 5(a)–5(c), we used the following rescaled parameters: $X_{\text{eq}} = 10.0$, $X_{\text{ad}} = 10.0$, and $\omega = 0.01$; in Case B, $X_{\text{eq}} = 10.0$, $X_{\text{ad}} = 50.0$, and $\omega = 10.0$. The numbers of the Monte Carlo steps are over 10^7

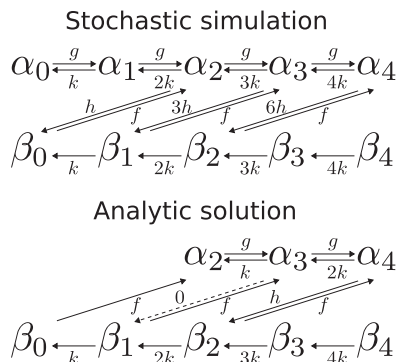


FIG. 4. Transition scheme for simulations and analytical calculations for dimer interaction cases.

for Case A and 10^8 for Case B. For Figs. 5(d) and 5(e), we set $X_{\text{eq}} = 10.0$.

In the dimer interaction cases, there are large differences in the variances of the number of proteins; see Fig. 5(c). This is because the crude approximation, i.e., the effective interaction \tilde{h} , does not include any fluctuation effects. However, the behavior of the modality in the approximation scheme is qualitatively similar to those of the Monte Carlo simulations.

IV. EXCLUSIVE SWITCH

Next, we consider a more complicated case, i.e., an exclusive switch [21–23]; see Fig. 6. The exclusive switch consists of two genes, gene 1 and gene 2. The two genes have overlapping promoter sites, and the binding of one of the regulatory proteins prevents the binding of the other regulatory proteins.

Here the interactions between the binding sites and proteins are assumed to be monomer interactions. Because the interactions are not cooperative bindings, the macroscopic rate equations give only one solution [21–23]. However, it has been shown that the exclusive switch can play as a switch. In the exclusive switch, stochastic effects make the bistability even without cooperativity between the regulatory proteins.

In order to study the exclusive switch analytically, master equations for a joint probability $P(n_1, n_2, s_1, s_2)$ should be constructed; n_i is the number of proteins for gene i , and $s_i \in \{\text{ON}, \text{OFF}\}$ indicates the gene state. In general, the master equations for multiple gene cases are very complicated, and it could be difficult to obtain numerical solutions if the number of genes is large.

In our approximation scheme, the effective interaction is used, and it enables us to reduce the dimensionality of the problem. Because we consider only the effective interaction, gene 1 and gene 2 can be considered independently, and then the probability $P(n_1, n_2, s_1, s_2)$ is expressed as $P(n_1, s_1) \times P(n_2, s_2)$. $P(n_i, s_i)$ is directly connected to the probabilities discussed in Sec. III B. For example, $P(n_1 + 1, \text{ON}) = P(\alpha_n^{(1)})$ and $P(n_1, \text{OFF}) = P(\beta_n^{(1)})$, where the superscript “(1)” indicates gene 1, and $\alpha_n^{(1)}$ and $\beta_n^{(1)}$ are states where there are n free regulatory proteins in an ON state and OFF state, respectively, for gene 1. Using these notations for the states, the master equations for gene 1 are written as

$$\begin{aligned} \frac{dP(\alpha_n^{(1)}, t)}{dt} &= g^{(1)}[P(\alpha_{n-1}^{(1)}, t) - P(\alpha_n^{(1)}, t)] \\ &\quad + k^{(1)}[(n+1)P(\alpha_{n+1}^{(1)}, t) - nP(\alpha_n^{(1)}, t)] \\ &\quad - \tilde{h}^{(1)}P(\alpha_n^{(1)}, t) + f^{(1)}P(\beta_n^{(1)}, t), \end{aligned} \quad (21)$$

$$\begin{aligned} \frac{dP(\beta_n^{(1)}, t)}{dt} &= k^{(1)}[(n+1)P(\beta_{n+1}^{(1)}, t) - nP(\beta_n^{(1)}, t)] \\ &\quad + \tilde{h}^{(1)}P(\alpha_n^{(1)}, t) - f^{(1)}P(\beta_n^{(1)}, t). \end{aligned} \quad (22)$$

The master equations for gene 2 can be obtained in a similar way. Using the same discussion for the self-regulating systems in Sec. III, the generating functions for genes $i \in \{1, 2\}$, $\alpha^{(i)}(z)$, and $\beta^{(i)}(z)$ are derived immediately.

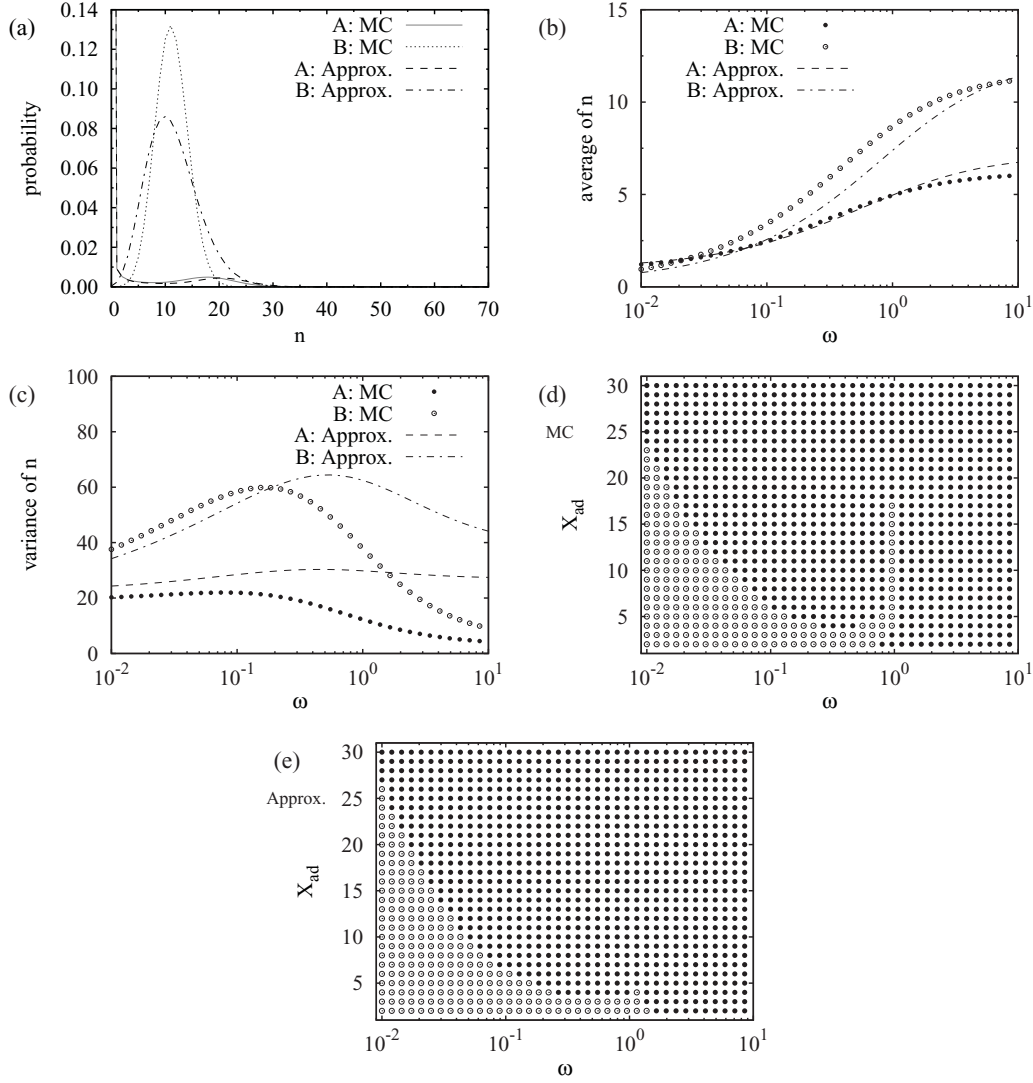


FIG. 5. Comparison between results from the Monte Carlo simulations and those from the approximation scheme. (a) Probability distributions of the number of proteins. $X_{eq} = 10.0$, $X_{ad} = 10.0$, $\omega = 0.01$ in Case A; $X_{eq} = 10.0$, $X_{ad} = 50.0$, $\omega = 10.0$ in Case B. (b) and (c) Averages and variances of the number of proteins, respectively. Only the value of ω was changed, and $X_{eq} = 10.0$, $X_{ad} = 10.0$ in Case A; $X_{eq} = 10.0$, $X_{ad} = 50.0$ in Case B. (d) and (e) Results for modality in the Monte Carlo simulations and in the approximation scheme, respectively. The filled circle means unimodal regions, and the empty circles indicate regions with bimodal distributions. $X_{eq} = 10.0$ for (d) and (e).

The evaluation of the effective interaction is slightly different from the self-regulating systems. The transition of gene 1 from an ON state to OFF state can occur only when gene 2 is in an ON state, and the effective interaction $\tilde{h}^{(1)}$

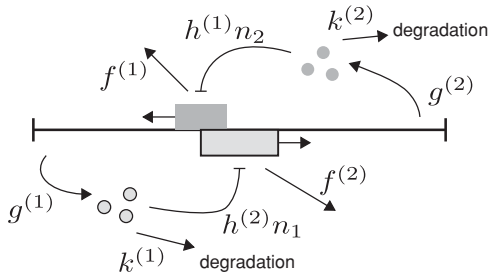


FIG. 6. A schematic illustration of the executive switch.

includes only a contribution from free proteins 2 in an ON state. As denoted above, gene 1 and gene 2 are independent, and hence gene 1 does not know whether gene 2 is in an ON state or OFF state; differently from Eq. (11), the conditional expectation cannot be used. Hence, the effective interaction for gene 1 is evaluated as

$$\tilde{h}^{(1)} = h^{(1)} \left. \frac{\partial}{\partial z} \alpha^{(2)}(z) \right|_{z=1}. \quad (23)$$

According to the above discussions, we finally obtain the following self-consistent equations:

$$\tilde{h}^{(1)} = \frac{h^{(1)} g^{(2)} f^{(2)} (k^{(2)} + f^{(2)})}{k^{(2)} (f^{(2)} + \tilde{h}^{(2)}) (k^{(2)} + f^{(2)} + \tilde{h}^{(2)})}, \quad (24)$$

$$\tilde{h}^{(2)} = \frac{h^{(2)} g^{(1)} f^{(1)} (k^{(1)} + f^{(1)})}{k^{(1)} (f^{(1)} + \tilde{h}^{(1)}) (k^{(1)} + f^{(1)} + \tilde{h}^{(1)})}. \quad (25)$$

By solving the above self-consistent equations, we obtain $\tilde{h}^{(1)}$ and $\tilde{h}^{(2)}$. We here solved them numerically.

Similar to Sec. III B, we can immediately obtain the probability distributions for each gene using the approximation scheme. In order to reconstruct the joint probability distribution for genes 1 and 2, we need more calculations as follows. First, we denote conditional probabilities for the number of free proteins for gene i ($i \in \{1,2\}$) as

$$P^{\text{cond}}(\alpha_n^{(i)}) \equiv \frac{P(\alpha_n^{(i)})}{\alpha^{(i)}(1)}, \quad (26)$$

$$P^{\text{cond}}(\beta_n^{(i)}) \equiv \frac{P(\beta_n^{(i)})}{\beta^{(i)}(1)}. \quad (27)$$

Second, because of the independence of genes, the joint probability distributions should be evaluated as

$$P(n_1, n_2, \text{ON}, \text{ON}) = P(\text{ON}, \text{ON}) P^{\text{cond}}(\alpha_{n_1-1}^{(1)}) P^{\text{cond}}(\alpha_{n_2-1}^{(2)}), \quad (28)$$

$$P(n_1, n_2, \text{ON}, \text{OFF}) = P(\text{ON}, \text{OFF}) P^{\text{cond}}(\alpha_{n_1-1}^{(1)}) P^{\text{cond}}(\beta_{n_2}^{(2)}), \quad (29)$$

$$P(n_1, n_2, \text{OFF}, \text{ON}) = P(\text{OFF}, \text{ON}) P^{\text{cond}}(\beta_{n_1}^{(1)}) P^{\text{cond}}(\alpha_{n_2-1}^{(2)}), \quad (30)$$

where $P(\text{ON}, \text{ON})$ is the probability with which gene 1 is in ON state and gene 2 is in ON state, and so on. Note that

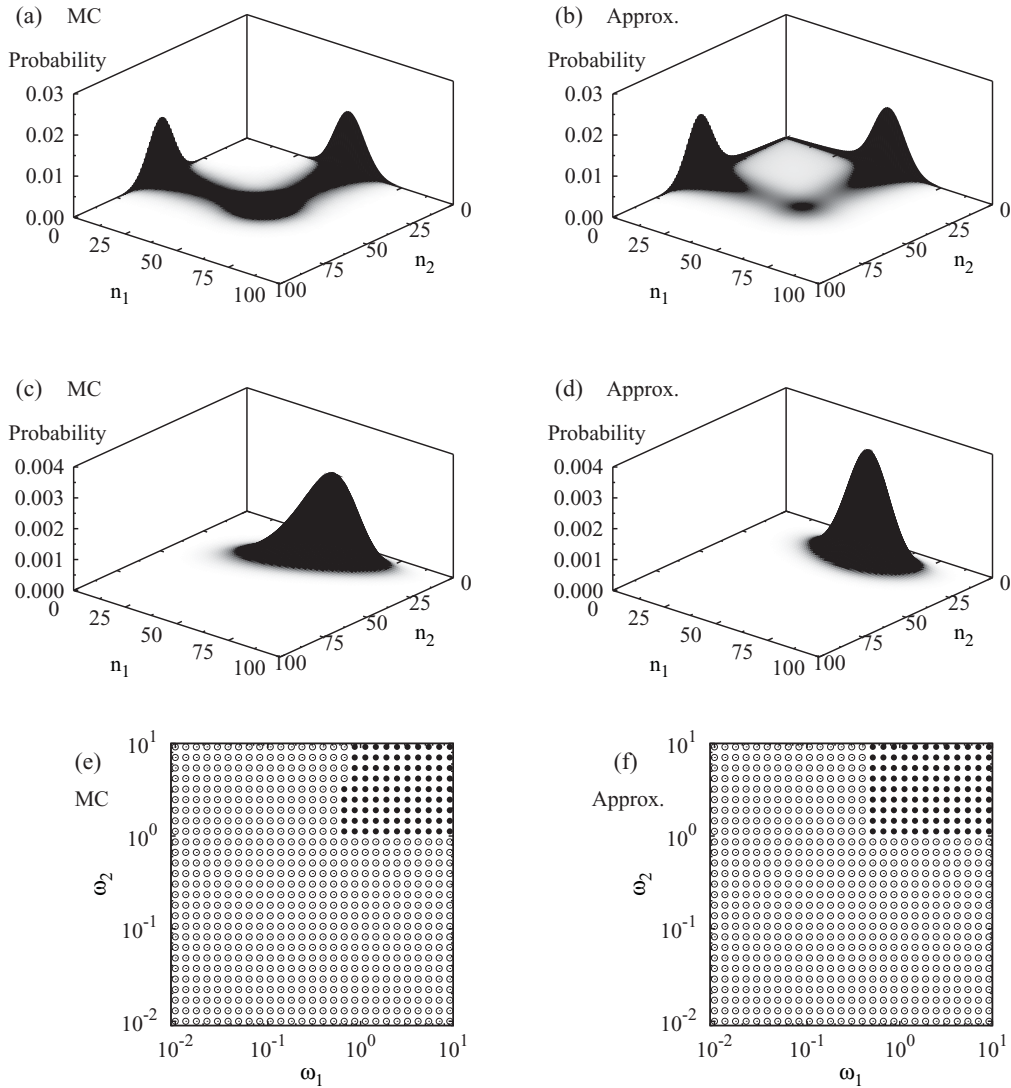


FIG. 7. Comparison between results from the Monte Carlo simulations and those from the approximation scheme for the exclusive switch. (a)–(d) Joint probability distributions. (a) Monte Carlo result for $X_{\text{eq}}^{(1)} = X_{\text{eq}}^{(2)} = 25.0$, $X_{\text{ad}}^{(1)} = X_{\text{ad}}^{(2)} = 25.0$, and $\omega^{(1)} = \omega^{(2)} = 0.1$; (b) the approximate results, and all parameters are the same as (a). (c) Monte Carlo results for $X_{\text{eq}}^{(1)} = 40.0$, $X_{\text{eq}}^{(2)} = 30.0$, $X_{\text{ad}}^{(1)} = 30.0$, $X_{\text{ad}}^{(2)} = 20.0$, $\omega^{(1)} = 10.0$ and $\omega^{(2)} = 20.0$. (d) Approximate results, and all parameters are the same as (c). (e) and (f) Results for modality in the Monte Carlo simulations and in the approximation scheme, respectively. The filled circle means unimodal regions, and the empty circles indicate regions with multimodal distributions. $X_{\text{eq}}^{(1)} = 20.0$, $X_{\text{ad}}^{(1)} = 10.0$, $X_{\text{eq}}^{(2)} = 30.0$, and $X_{\text{ad}}^{(2)} = 20.0$ for (e) and (f).

$\alpha_{n_i-1}^{(i)}$ means only the number of “free” proteins; for monomer interaction cases, the difference between $\alpha_n^{(i)}$ and $\alpha_n^{(i)}$ is only one inert protein, and there is no such difference for OFF state. In addition, the probability with which both genes 1 and 2 are in OFF state is zero, $P(\text{OFF}, \text{OFF}) = 0$, because of the exclusive settings.

Taking the exclusive settings into account, the marginal probabilities are calculated as follows:

$$\begin{cases} P(1_{\text{ON}}) = P(\text{ON}, \text{ON}) + P(\text{ON}, \text{OFF}), \\ P(1_{\text{OFF}}) = P(\text{OFF}, \text{ON}), \\ P(2_{\text{ON}}) = P(\text{ON}, \text{ON}) + P(\text{OFF}, \text{ON}), \\ P(2_{\text{OFF}}) = P(\text{ON}, \text{OFF}), \end{cases} \quad (31)$$

where $P(1_{\text{ON}})$ is a probability with which gene 1 is in ON state, and so on. Hence, we obtain

$$\begin{cases} P(\text{ON}, \text{ON}) = P(1_{\text{ON}}) - P(2_{\text{OFF}}), \\ P(\text{ON}, \text{OFF}) = P(2_{\text{OFF}}), \\ P(\text{OFF}, \text{ON}) = P(1_{\text{OFF}}), \\ P(\text{OFF}, \text{OFF}) = 0. \end{cases} \quad (32)$$

The marginal probabilities, such as $P(1_{\text{ON}})$, can be evaluated by using the generating functions for each gene. Finally, we can construct the joint probability distribution as

$$\begin{aligned} P(n_1, n_2) &= P(n_1, n_2, \text{ON}, \text{ON}) + P(n_1, n_2, \text{OFF}, \text{ON}) \\ &\quad + P(n_1, n_2, \text{ON}, \text{OFF}). \end{aligned} \quad (33)$$

We here note that the probabilities $P(\text{ON}, \text{ON})$, calculated using the above procedures, may become negative for some cases; i.e., $P(1_{\text{ON}}) > P(2_{\text{OFF}})$ for some choices of parameters $g^{(i)}, k^{(i)}, h^{(i)}$, and $f^{(i)}$. This is because we neglect correlations between two genes. In these cases, other procedures to estimate the joint probabilities are needed. We have not yet succeeded in finding general conditions to cause the negative joint probabilities, and in the following numerical experiments, only the former cases [$P(1_{\text{ON}}) < P(2_{\text{OFF}})$] are treated.

Figures 7(a)–(d) show the joint probability distributions. Figures 7(a) and 7(c) are Monte Carlo results, and Figs. 7(b) and 7(d) are results of the approximation scheme. As in Sec. III C, we used the rescaled parameters and set $k^{(1)} = k^{(2)} = 1$. In Figs. 7(a) and 7(b), we used the parameters $X_{\text{eq}}^{(1)} = X_{\text{eq}}^{(2)} = 25.0$, $X_{\text{ad}}^{(1)} = X_{\text{ad}}^{(2)} = 25.0$, and $\omega^{(1)} = \omega^{(2)} = 0.1$; for 7(c) and 7(d), $X_{\text{eq}}^{(1)} = 40.0$, $X_{\text{eq}}^{(2)} = 30.0$, $X_{\text{ad}}^{(1)} = 30.0$, $X_{\text{ad}}^{(2)} = 20.0$, $\omega^{(1)} = 10.0$, and $\omega^{(2)} = 20.0$. The numbers of the Monte Carlo steps are over 10^8 for Fig. 7(a), and over 10^9 for Fig. 7(c). Although Fig. 7(d) does not show the correlated behavior seen in Fig. 7(c) because correlations between gene 1 and gene 2 are largely neglected in the approximation scheme, one could say that the approximation scheme gives qualitatively good results; the characteristics of the peak structure are recovered adequately despite the rough approximation. In Fig. 7(b), the bistability due to the exclusive settings is recovered well.

Finally, we investigate the modality in the exclusive settings. For the two-gene case, we define the modality as follows: If both the marginal distributions for n_1 and n_2 are unimodal, we define the joint probability distribution

as unimodal; otherwise, the joint probability distribution is defined as multimodal. Figures 7(e) and 7(f) show the results for the Monte Carlo simulations and for the approximation scheme, respectively. Here we set $X_{\text{eq}}^{(1)} = 20.0$, $X_{\text{ad}}^{(1)} = 10.0$, $X_{\text{eq}}^{(2)} = 30.0$, and $X_{\text{ad}}^{(2)} = 20.0$. We note that in the parameter region in Fig. 7(f), we do not have any negative joint probabilities. From Figs. 7(e) and 7(f), we can see that the approximation scheme qualitatively recovers the behavior of the modality.

V. CONCLUDING REMARKS

In the present paper, we developed the approximation scheme for gene regulatory systems. We first applied it to self-regulating systems. The approximation scheme gives qualitatively good results; the characteristics of peak structures can be recovered well. In addition, due to the extension of the basic idea of the effective interactions, we can naturally apply the approximation scheme even to the exclusive switch, and the bistability of the exclusive switch without cooperative interactions is successfully recovered.

In contrast to the Fokker-Planck or Langevin approach, the approximation scheme proposed in the present paper does not neglect discrete properties of systems. In addition, because we can rewrite the joint probability for all genes as a product of the probability distribution for each gene, the dimensions of the problems are reduced to a large degree.

Although a switching time between two stable states is also an important quantity in bistable systems such as genetic switches, the evaluation of the switching time is outside the scope of the current work. In the present paper, we focus only on “static” information, such as the average number of proteins and probability distributions for protein numbers. Approximation schemes even for static information has been needed and actually developed. For example, in transcriptional regulatory cascades, information theoretic analysis has been performed [18]; mutual information for long cascading systems should be evaluated in the information theoretic analysis, but Monte Carlo simulations are sometimes time consuming and need high computational costs, especially for the long cascading systems. In order to calculate the mutual information for such systems, a spectral analysis method has been developed, which treats static information [26]. We expect that our approximation scheme should help future work related to information theoretic studies. In addition, it is important to study cases in which equations for the effective interactions have bifurcation points. For example, a self-activator system causes the bifurcation problem. It is not straightforward to construct a whole probability distribution from two solutions for the effective interaction, and these cases are under investigation.

ACKNOWLEDGMENTS

The author thanks Masaki Sasai for helpful comments on this manuscript. This work was supported in part by a grant-in-aid for scientific research (Nos. 20115009 and 21740283) from the Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan.

- [1] C. V. Rao, D. M. Wolf, and A. P. Arkin, *Nature (London)* **420**, 231 (2002).
- [2] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, *Science* **297**, 1183 (2002).
- [3] M. Kærn, T. C. Elston, W. J. Blake, and J. J. Collins, *Nat. Rev. Genet.* **6**, 451 (2005).
- [4] T. S. Gardner, C. R. Cantor, and J. J. Collins, *Nature (London)* **403**, 339 (2000).
- [5] H. Okano, T. J. Kobayashi, H. Tozaki, and H. Kimura, *Biophys. J.* **95**, 1063 (2008).
- [6] D. Schultz, J. N. Onuchic, and P. G. Wolynes, *J. Chem. Phys.* **126**, 245102 (2007).
- [7] M. Yoda, T. Ushikubo, W. Inoue, and M. Sasai, *J. Chem. Phys.* **126**, 115101 (2007).
- [8] J. Hasty, J. Pradines, M. Dolnik, and J. J. Collins, *Proc. Natl. Acad. Sci. USA* **97**, 2075 (2000).
- [9] T. B. Kepler and T. C. Elston, *Biophys. J.* **81**, 3116 (2001).
- [10] W. Bialek, in *Advances in Neural Information Processing Systems*, edited by T. K. Leen, T. G. Dietterich, and V. Tresp (MIT Press, Cambridge, 2001), Vol. 13.
- [11] M. Sasai and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **100**, 2374 (2003).
- [12] J. E. M. Hornos, D. Schultz, G. C. P. Innocentini, J. Wang, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, *Phys. Rev. E* **72**, 051907 (2005).
- [13] A. M. Walczak, M. Sasai, and P. G. Wolynes, *Biophys. J.* **88**, 828 (2005).
- [14] K.-Y. Kim and J. Wang, *PLoS Comp. Biol.* **3**, e60 (2007).
- [15] J. Ohkubo, *J. Stat. Mech.* (2007) P09017.
- [16] V. Shahrezaei and P. S. Swain, *Proc. Natl. Acad. Sci. USA* **105**, 17256 (2008).
- [17] P. Visco, R. J. Allen, and M. R. Evans, *Phys. Rev. E* **79**, 031923 (2009).
- [18] A. M. Walczak and P. G. Wolynes, *Biophys. J.* **96**, 4525 (2009).
- [19] J. L. Cherry and F. R. Adler, *J. Theor. Biol.* **203**, 117 (2000).
- [20] P. B. Warren and P. R. ten Wolde, *Phys. Rev. Lett.* **92**, 128101 (2004).
- [21] A. Lipshtat, A. Loinger, N. Q. Balaban, and O. Biham, *Phys. Rev. Lett.* **96**, 188101 (2006).
- [22] A. Loinger, A. Lipshtat, N. Q. Balaban, and O. Biham, *Phys. Rev. E* **75**, 021904 (2007).
- [23] D. Schultz, A. M. Walczak, J. N. Onuchic, and P. G. Wolynes, *Proc. Natl. Acad. Sci. USA* **105**, 19165 (2008).
- [24] H. Feng, B. Han, and J. Wang, *J. Phys. Chem. B* **115**, 1254 (2011).
- [25] D. T. Gillespie, *J. Phys. Chem.* **81**, 2340 (1977).
- [26] A. M. Walczak, A. Mugler, and C. H. Wiggins, *Proc. Natl. Acad. Sci. USA* **106**, 6529 (2009).