

# Potential Energy Landscape and Robustness of a Gene Regulatory Network: Toggle Switch

Keun-Young Kim<sup>1</sup>, Jin Wang<sup>1,2,3\*</sup>

**1** Department of Physics and Astronomy, State University of New York Stony Brook, Stony Brook, New York, United States of America, **2** Department of Chemistry, State University of New York Stony Brook, Stony Brook, New York, United States of America, **3** State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin, People's Republic of China

**Finding a multidimensional potential landscape is the key for addressing important global issues, such as the robustness of cellular networks. We have uncovered the underlying potential energy landscape of a simple gene regulatory network: a toggle switch. This was realized by explicitly constructing the steady state probability of the gene switch in the protein concentration space in the presence of the intrinsic statistical fluctuations due to the small number of proteins in the cell. We explored the global phase space for the system. We found that the protein synthesis rate and the unbinding rate of proteins to the gene were small relative to the protein degradation rate; the gene switch is monostable with only one stable basin of attraction. When both the protein synthesis rate and the unbinding rate of proteins to the gene are large compared with the protein degradation rate, two global basins of attraction emerge for a toggle switch. These basins correspond to the biologically stable functional states. The potential energy barrier between the two basins determines the time scale of conversion from one to the other. We found as the protein synthesis rate and protein unbinding rate to the gene relative to the protein degradation rate became larger, the potential energy barrier became larger. This also corresponded to systems with less noise or the fluctuations on the protein numbers. It leads to the robustness of the biological basins of the gene switches. The technique used here is general and can be applied to explore the potential energy landscape of the gene networks.**

Citation: Kim KY, Wang J (2007) Potential energy landscape and robustness of a gene regulatory network: Toggle switch. *PLoS Comput Biol* 3(3): e60. doi:10.1371/journal.pcbi.0030060

## Introduction

In the post-genome era, with a wealth of data on genomic sequences, the crucial question becomes how to understand the organization of these sequences in nature and how genes function [1–4]. This is a challenging task. According to the central dogma, turning gene switches on and off controls certain types of protein synthesis and production. Furthermore, the on and off of gene switches determines the developmental plans of the cell. On the other hand, the protein products generated by the gene switches act back on the genes to control their expression patterns. The gene regulations thereby form a network with inherent many-body interactions and feedback loops. That is why the system often becomes quite complicated and hard to study.

The underlying nature of cellular networks has been explored by many experimental techniques [4]. It has often been found that cellular networks are in general quite robust and perform their biological functions in the midst of environmental perturbations. There have recently been an increasing number of studies on the global topological structures of cellular networks [5–8]. However, so far there are very few studies from the physical point of view of why the networks are so robust and why they perform their biological functions [9–20].

Theoretical models of cellular networks have often been formulated with a set of chemical reaction equations in bulk. These averaged dynamical descriptions are inherently local. To probe the global properties, one often has to explore different parameters. Since the parameter space is huge, the issue of global robustness is hard to address directly from these approaches.

Here we will explore the nature of the network from another angle: we formulate the problem in terms of the potential energy function or potential energy landscape. If the potential landscape of the cellular network is known, the global properties can be explored [21,22]. This is analogous to the fact that the global thermodynamic properties can be explored when knowing the inherent interaction potentials in a system.

There is another intriguing factor controlling the gene expression patterns. In the cell, there are a finite number of molecules (typically on the order of several hundreds or thousands). The intrinsic statistical fluctuations, usually not encountered in bulk due to the large-number averaging, can be significant and play an important role in the dynamics of gene expression. This gives the source of intrinsic statistical fluctuations or noise. On the other hand, the fluctuations from highly dynamical and nonhomogeneous environments of the interior of the cell give the source of the external noise for the networks [23–30]. It is important to investigate the roles of the statistical fluctuations or noises on the robustness and stability of the network.

In general, instead of studying the averaged chemical

**Editor:** Wei Wang, University of North Carolina, United States of America

**Received:** September 25, 2006; **Accepted:** February 14, 2007; **Published:** March 30, 2007

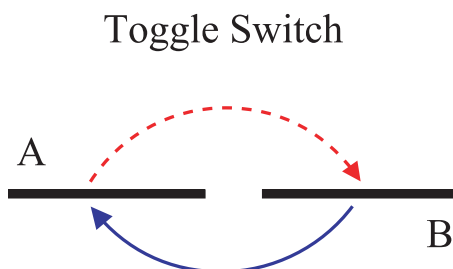
**Copyright:** © 2007 Kim and Wang. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

\* To whom correspondence should be addressed. E-mail: jin.wang.1@stonybrook.edu

## Author Summary

Cellular networks are at the heart of systems biology at present. To understand how cellular networks function in these highly fluctuating environments, a global approach is needed. Here we provide a global framework, in terms of potential landscapes, for studying the gene regulatory networks in the presence of the intrinsic statistical fluctuations. We uncovered the underlying landscape for the network. We identified the basins of attraction of the landscape as the biological functional states. The potential barrier between the two basins determines the time scale of conversion from one to the other. The robustness of the biological functional states of the network, the gene switches in this case, can be guaranteed if the conversions among the basins of attraction are not frequent, or, in other words, the barriers among the basins are relatively large. More detailed features of the network, such as the key genes or regulating links relevant to diseases (i.e., cancers), can be uncovered from the underlying landscape. Our technique is general and can be applied to explore the potential landscape of more realistic gene networks. Furthermore, our approach can also be helpful in guiding the network optimal design for synthetic biology.

reaction network equations in bulk, we should use statistical descriptions to model the cellular process. This can be realized by constructing a master equation for the evolution of probability instead of average concentration for the corresponding chemical reaction network equations [26,31–35]. One can also study the steady state properties of these probabilistic chemical reaction network equations. The generalized potential energy for the steady state of the network can be shown to be closely associated with the steady state probability of the network in general [10–12,15–20,31,32]. Once the network problem has been formulated in terms of the generalized potential function or potential landscape, the issue of the global stability or robustness is much easier to address. In fact, some explicit illustrations of the potential energy landscape and robustness for the MAP Kinase signal transduction network and cell cycle have been given recently [19,20]. It is the purpose of this paper to study the global robustness problem directly from the properties of the potential landscape for a simple yet important gene regulatory network: a toggle switch. Figure 1 shows a toggle switch. Gene networks often involve many degrees of freedom. To resolve the issue of multidimensionality, instead of using the direct Monte Carlo simulation [33] for solving the master equations, a Hartree mean field approximation can be applied to reduce the dimensionality and address the global issues [11,12,36–38].



**Figure 1.** Toggle Switch: Protein A Represses Gene B (Dotted Line), and Protein B Represses Gene A (Solid Line)  
doi:10.1371/journal.pcbi.0030060.g001

There are three aims of this paper. Our first aim is to develop a time-dependent Hartree approximation scheme [36] to solve the associated master equations to follow the evolution of multidimensional probability of the network. Our second aim is to construct the underlying potential energy landscape for a toggle switch [39] and explore both the steady state and time evolution of the landscape. Our third aim is to study the phase diagram of the system and the kinetic time scale from one stable basin of attraction to another in different conditions. We will address the global robustness condition for a toggle switch.

## Methods and Materials

As our goal is to uncover the potential energy landscape, we first studied the chemical reaction network involved in gene regulations. In particular we need to take into account the intrinsic statistical fluctuations due to the finite number of molecules in the cells. The statistical nature of the chemical reactions can be captured by the corresponding master equations. We established master equations for the gene regulations that describe the evolution of the networks probabilistically. The master equation is almost impossible to solve due to its inherent huge dimensions. We therefore used the Hartree approximation to reduce the dimensionality [11,12,36]. In this way, we could follow the time evolution and steady state probability of the protein concentrations. The steady state probability is closely associated with the underlying potential energy landscape, which is our ultimate target.

## Assumptions

Gene expression is regulated in various and complex ways, and can be represented by many coupled biochemical reactions. In this report, our goal was not just to explain some specific gene network system as accurately as possible, but to illustrate mathematical tools for exploring the general mechanisms of transcriptional regulatory gene networks. We therefore took abstractions of some essential biochemical reactions from complicated reactions of diverse systems.

Let us start with the explanation of some terminologies used in this manuscript: “activator” is a regulatory protein that increases the level of transcription, “repressor” is a regulatory protein that decreases the level of transcription. By “operator” we mean the DNA site or the gene where regulatory proteins (either an activator or a repressor) bind. First we are interested in the effect of “operator fluctuation” by which we mean the biochemical reactions that change the state of the operator. The operator is said to be in an occupied state if a regulatory protein is bound to it, and in an unoccupied state if the protein is not bound to it. For the repressor we include the following reaction.



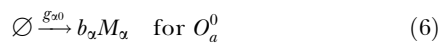
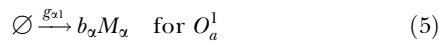
where  $O_{\alpha}^{1(0)}$  stands for the active (inactive) operator state of gene  $\alpha$ ,  $M_{\beta}$  represents the regulatory protein synthesized or produced by gene  $\beta$ , and  $q_{\alpha\beta}$  is for the multimer-type of proteins. For example, if  $q_{AB} = 2(3)$ , dimer (tetramer) proteins produced from gene B repress the expression of gene A,  $h_{\alpha\beta}^R$

and  $f_{\alpha\beta}^R$  are reaction probabilities per unit time. In a similar way, we may also consider the activator:



Notice that the superscript 1(0) in  $O_{\alpha}^{1(0)}$  indicates the activity state of the operator and does not represent the bound state of regulatory protein. We will say the gene is on (off) when the operator of the gene is active (inactive). The gene will be on when it is occupied by activators or when repressors are unbound from it.

Next we include the transcription and translation steps. Here we ignore mRNA and consider only one step combining transcription and translation:



where  $\emptyset$  denotes a protein sink or source,  $b_{\alpha}$  stands for the burst size of produced proteins ( $M_{\alpha}$ ),  $g_{1(0)}$  is a protein synthesis probability per unit time, and  $k_{\alpha}$  is the degradation probability per unit time.

We can say that Equations 1–7 are “effective reactions” of the transcriptional regulatory gene network system. Roughly speaking, we can say the other biochemical reactions could be taken into account by adjusting the parameters of the reaction probabilities per unit time. In this sense, the reaction parameters are not really constants but functions of time. Furthermore, the proteins may not be well-mixed in the cell, and the number of proteins could be a function of position. So we can generalize this formalism in a space-dependent manner. We also can add more species and reactions to the master equations. In this report we will assume homogeneity of the number of proteins and ignore the time delay (for example, due to the translation process) so that all the parameters are constants. Now we can construct the master equation based on the above assumptions and chosen effective reactions.

### Hartree Approximation of the Master Equation

The master equation is the equation for the time evolution of the probability of some specific state  $P$ :

$$P(n_A, S_A, n_B, S_B, n_C, S_C, \dots, t) \quad (8)$$

where A,B,C, ... is the label of each gene;  $n_A, n_B, n_C, \dots$  is the number of proteins expressed by gene A,B,C, ... , respectively.  $S_A, S_B, S_C, \dots$  is 1 or 0, and represents the activity state of the operator. The number of states,  $N$ , is  $n_A \times 2 \times n_B \times 2 \times n_C \times 2 \times \dots$ . We expected to have  $N$ -coupled differential equations, which are not feasible to solve. Following a mean field approach [11], we used the Hartree approximation to split the probability into the products of individual ones: First, let us assume

$$P(n_A, S_A, n_B, S_B, n_C, S_C, \dots, t) = \prod_{\beta} P(n_{\beta}, S_{\beta}, t), \quad (9)$$

and sum over all indexes except one specific index that we are interested in, say  $\alpha$ . This effectively reduces the dimensionality from exponential  $n_A \times n_B \times \dots \times n_N \times 2^N$  to multiples  $(n_A + n_B + \dots) \times 2 \times N$ , and therefore the problem is computationally tractable. Finally we are left with two equations, one for  $P(n_{\alpha}, 1)$  and one for  $P(n_{\alpha}, 0)$ . (In fact, these are not just two equations because  $n_{\alpha}$  varies from 0 to hundreds.) With the two component vector notation,

$$\mathbf{P}_{\alpha} := \begin{pmatrix} P_{\alpha 1}(n_{\alpha}, t) \\ P_{\alpha 0}(n_{\alpha}, t) \end{pmatrix} := \begin{pmatrix} P(n_{\alpha}, S_{\alpha} = 1, t) \\ P(n_{\alpha}, S_{\alpha} = 0, t) \end{pmatrix} \quad (10)$$

we have the compact form for the network:

$$\begin{aligned} \frac{\partial}{\partial t} \mathbf{P}_{\alpha}(n_{\alpha}, t) &= \begin{pmatrix} g_{\alpha 1} & 0 \\ 0 & g_{\alpha 0} \end{pmatrix} \\ &\times \{ \theta(n_{\alpha} - b_{\alpha}) \mathbf{P}_{\alpha}(n_{\alpha} - b_{\alpha}, t) - \mathbf{P}_{\alpha}(n_{\alpha}, t) \} \\ &+ k_{\alpha} \{ (n_{\alpha} + 1) \mathbf{P}_{\alpha}(n_{\alpha} + 1, t) - n_{\alpha} \mathbf{P}_{\alpha}(n_{\alpha}, t) \} \\ &+ \left[ \underbrace{\sum_{\beta} \begin{pmatrix} -H_{\alpha\beta}^R & f_{\alpha\beta}^R \\ H_{\alpha\beta}^R & -f_{\alpha\beta}^R \end{pmatrix}}_{\text{Repressive interaction}} \right. \\ &\quad \left. + \underbrace{\sum_{\beta} \begin{pmatrix} -f_{\alpha\beta}^P & H_{\alpha\beta}^P \\ f_{\alpha\beta}^P & -H_{\alpha\beta}^P \end{pmatrix}}_{\text{Promoting interaction}} \right] \mathbf{P}_{\alpha}(n_{\alpha}, t) \quad (11) \end{aligned}$$

where

$$\begin{aligned} H_{\alpha\beta}^{R(P)} &:= h_{\alpha\beta}^{R(P)} \left\langle \frac{n_{\beta}!}{(n_{\beta} - q_{\alpha\beta})! q_{\alpha\beta}!} \right\rangle, \\ &=_{n_{\beta} \mathbf{C}_{q_{\alpha\beta}}} \\ q_{\alpha\beta} &:= \begin{pmatrix} 1 & \text{monomer} \\ 2 & \text{dimer} \\ 3 & \text{trimer} \\ q & q\text{-th multimer} \end{pmatrix}. \quad (12) \end{aligned}$$

Notice that Equation 11 is simply a “birth–death” process without the last term. We will call the first two terms in Equation 11 the birth–death part or “drift and diffusion” part from the viewpoint of the diffusional Fokker–Plank equation [10,35]. Furthermore, we will call the last term the operator fluctuation part. In Equation 11, all other indices except  $\alpha$  appear only in  $H_{\alpha\beta}$  in the ensemble-averaged form ( $f_{\alpha\beta}$  is just some number). If we deal with the one gene case, there is no ensemble average in Equation 12. The first effect of the operator fluctuation is the sum over  $n_{\beta}$  and  $S_{\beta}$ . The second effect is to cancel out many of the birth–death terms of other genes. Since  $\alpha = A, B, C, \dots$ , we have the vector equation set of the same numbers as those of the genes. They are coupled to each other through the term  $H_{\alpha\beta}$ . All network interactions can be determined by assigning every  $h_{\alpha\beta}$ .  $b_{\alpha}$  is the number of proteins produced in bursts from gene  $\alpha$ , and  $\theta$  is a step function. In Equation 12, we take into account several kinds of binding proteins, and use proper combinatorics and ensemble average.

## Quantum Field Theoretic Description

The techniques of quantum field theory can be used to solve the master equation [11,37]. The first step is to construct a many-body quantum state. Notice that the probabilities defined by Equation 10 are imbedded in the quantum state as coefficients (Equation 14)

$$|\Psi\rangle := \prod_{\alpha} |\psi_{\alpha}\rangle = |\psi_A\rangle \otimes |\psi_B\rangle \otimes |\psi_C\rangle \otimes \dots \quad (13)$$

where

$$|\psi_{\alpha}\rangle := \left( \sum_{n_{\alpha}} P_{\alpha 1}(n_{\alpha}, t) |n_{\alpha}\rangle \right) \quad (14)$$

In Equation 13 we make an ansatz of Hartree-type product for the many-body state. Then non-Hermitian ‘‘Hamiltonian’’ of only repressive proteins,  $\Omega$ , yields:

$$\Omega := \sum_{\alpha} \left( \left( \mathbf{D}_{\alpha} + \underbrace{\begin{pmatrix} -\mathbf{H}_{\alpha\alpha} & 0 \\ \mathbf{H}_{\alpha\alpha} & 0 \end{pmatrix}}_{\text{self repressive}} \right) \otimes \prod_{\beta \neq \alpha} 1_{\beta} \right) \quad (15)$$

$$+ 1_{\alpha} \otimes \underbrace{\sum_{\beta \neq \alpha} \begin{pmatrix} -\mathbf{H}_{\alpha\beta} & 0 \\ \mathbf{H}_{\alpha\beta} & 0 \end{pmatrix} \otimes \prod_{\gamma \neq \alpha, \beta} 1_{\gamma}}_{\text{interaction part(network)}}$$

where

$$\mathbf{D}_{\alpha} := \begin{pmatrix} g_{\alpha 1}((a_{\alpha}^{+})^{b_{\alpha}} - 1) + k_{\alpha}(a_{\alpha} - a_{\alpha}^{+}a_{\alpha}) & f_{\alpha} \\ 0 & g_{\alpha 0}((a_{\alpha}^{+})^{b_{\alpha}} - 1) + k_{\alpha}(a_{\alpha} - a_{\alpha}^{+}a_{\alpha}) - f_{\alpha} \end{pmatrix} \quad (16)$$

$$\mathbf{H}_{\alpha\beta} := \sum_{q_{\alpha\beta}} h_{\alpha\beta} \frac{(a_{\beta}^{+}a_{\beta})!}{(a_{\beta}^{+}a_{\beta} - q_{\alpha\beta})!q_{\alpha\beta}!} \quad (17)$$

For each protein concentration, a creation and an annihilation operator are introduced, such that  $a^{+}|n\rangle = |n+1\rangle$  and  $a|n\rangle = n|n-1\rangle$ . These operators satisfy  $[a, a^{+}] = 1$ . The generalization to include activating proteins is straightforward. While the state vector is a simple product of individual genes, the operator product form of  $\Omega$  is chosen deliberately to reproduce the original master Equation 11. The  $\Omega$  of a many-gene system seems to be  $\Omega = \sum \Omega_i$  [11], but it would not be the simple sum of individual operators because of the interaction terms. Like the master equation,  $\mathbf{D}_{\alpha}$  is the birth-death part and plays a role in the diffusion and drift terms in the context of Fokker-Plank equation. The second term and third term in Equation 15 are repressor-related terms, and  $\mathbf{H}_{\alpha\beta}$  is the counterpart of  $M_{\alpha\beta}$  in Equation 12. Finally, we have the following quantum system:

$$\frac{\partial}{\partial t} |\Psi\rangle = \Omega |\Psi\rangle. \quad (18)$$

To complete the mean-field approximation, we need to average all interaction effects by doing an inner product with some reference state, which is a two-component generalization of the Glauber state [37]. If we are interested in an  $\alpha$ -gene (operator) state and the associated protein, we may define the reference state:

$$\langle \tilde{\Phi}_{\alpha} | := \prod_{\beta \neq \alpha} \langle S_{\beta} | := \prod_{\beta \neq \alpha} (\langle 0 | e^{a_{\beta}} \langle 0 | e^{a_{\beta}}). \quad (19)$$

Then,

$$\langle \tilde{\Phi}_{\alpha} | \frac{\partial}{\partial t} |\Psi\rangle = \langle \tilde{\Phi}_{\alpha} | \Omega |\Psi\rangle \quad (20)$$

is equivalent to the master Equation 11.

## Solutions

**Ritz’s variational method with coherent state ansatz.** We will use the Rayleigh–Ritz variational method to obtain an approximate solution of a non-Hermitian Hamiltonian system (nonequilibrium system) like Equation 18 [11,38]. The master equation is equivalent to the functional variation  $\delta\Gamma/\delta\Phi = 0$  of an effective action  $\Gamma = \int dt \langle \Phi | (\partial_t - \Omega) | \Psi \rangle$ . The method is analogous to the traditional procedure in quantum mechanics, except the modification due to non-Hermitian properties of the operator, of which left eigenvectors and right eigenvectors do not have to be the same. We will make a ket state ansatz,  $|\Psi\rangle$ , and a bra state ansatz,  $\langle \Phi |$ , respectively [11]:

$$|\Psi\rangle := \prod_{\alpha} |\psi_{\alpha}\rangle \quad \langle \Phi | := \prod_{\alpha} \langle \phi_{\alpha} | \quad (21)$$

$$|\psi_{\alpha}\rangle := \begin{pmatrix} C_{\alpha 1} e^{X_{\alpha 1} (a_{\alpha}^{+} - 1)} |0_{\alpha}\rangle \\ C_{\alpha 0} e^{X_{\alpha 0} (a_{\alpha}^{+} - 1)} |0_{\alpha}\rangle \end{pmatrix} \quad (22)$$

$$\langle \phi_{\alpha} | := (\langle 0_{\alpha} | e^{a_{\alpha}} e^{\sigma_{\alpha 1} + \lambda_{\alpha 1} a_{\alpha}} \quad \langle 0_{\alpha} | e^{a_{\alpha}} e^{\sigma_{\alpha 0} + \lambda_{\alpha 0} a_{\alpha}}) \quad (23)$$

where  $C_{\alpha 1}$ ,  $C_{\alpha 0}$ ,  $X_{\alpha 1}$ ,  $X_{\alpha 0}$ ,  $\alpha_{\alpha 1}$ ,  $\alpha_{\alpha 0}$ ,  $\lambda_{\alpha 1}$ ,  $\lambda_{\alpha 0}$  are time-dependent parameters to be determined by the variational principle. The ket ansatz is chosen as coherent state, which corresponds to a Poisson distribution.  $C_{\alpha 1}$  and  $C_{\alpha 0}$  are the probabilities of the two DNA-binding states,  $S = 1$  and  $S = 0$ , respectively. The coupled dynamics of the DNA-binding state and the protein distribution are described as the motion of wavepackets with amplitudes  $C_{\alpha 1}$  and  $C_{\alpha 0}$  as well as by means of the protein concentrations at  $X_{\alpha 1}$  and  $X_{\alpha 0}$  (from the Poisson distribution ansatz). With the following notation

$$\alpha_{\alpha j}^R = C_{\alpha 1}, C_{\alpha 0}, X_{\alpha 1}, X_{\alpha 0} \quad j = 1, 2, 3, 4, \quad (24)$$

respectively  $\alpha = A, B, C, \dots$

$$\alpha_{\alpha j}^L = \alpha_{\alpha 1}, \alpha_{\alpha 0}, \lambda_{\alpha 1}, \lambda_{\alpha 0} \quad j = 1, 2, 3, 4, \quad (25)$$

respectively  $\alpha = A, B, C, \dots$

Here,  $\langle \Phi | (\alpha^L = 0)$  is set to be consistent with the probabilistic interpretation  $\langle \Phi | (\alpha^L = 0) | \Psi (\alpha^R) \rangle = 1$ . The condition for the extremum of the action with respect to  $\langle \Phi |$  is:

$$\left\langle \frac{\partial \Psi}{\partial \alpha_{\alpha i}^L} \middle| \frac{\partial \Psi}{\partial \alpha_{\alpha j}^R} \right\rangle \frac{d\alpha_{\alpha j}^R}{dt} = \left\langle \frac{\partial \Psi}{\partial \alpha_{\alpha i}^L} \middle| \Omega | \Psi \right\rangle, \quad \text{all } \alpha_{\alpha i}^L = 0 \quad (26)$$

which is reduced to the coupled ordinary differential equations with parameters:

$$\frac{dC_{\alpha 1}}{dt} = -C_{\alpha 1}H_{total}^R + C_{\alpha 0}f_{total}^R + C_{\alpha 0}H_{total}^P - C_{\alpha 1}f_{total}^P \quad (27)$$

$$\frac{dC_{\alpha 0}}{dt} = -\frac{dC_{\alpha 1}}{dt} \quad (28)$$

$$C_{\alpha 1} \frac{dX_{\alpha 1}}{dt} + X_{\alpha 1} \frac{dC_{\alpha 1}}{dt} = C_{\alpha 1}(b_{\alpha}g_{\alpha 1} - k_{\alpha}X_{\alpha 1}) - C_{\alpha 1}H_{total}^R + C_{\alpha 0}f_{total}^R X_{\alpha 0} + C_{\alpha 0}H_{total}^P - f_{total}^P C_{\alpha 1} X_{\alpha 1} \quad (29)$$

$$C_{\alpha 0} \frac{dX_{\alpha 0}}{dt} + X_{\alpha 0} \frac{dC_{\alpha 0}}{dt} = C_{\alpha 0}(b_{\alpha}g_{\alpha 0} - k_{\alpha}X_{\alpha 0}) + C_{\alpha 1}H_{total}^R - C_{\alpha 0}f_{total}^R X_{\alpha 0} - C_{\alpha 0}H_{total}^P + f_{total}^P C_{\alpha 1} X_{\alpha 1} \quad (30)$$

$$H_{total}^R := \langle 0 | e^{a_{\alpha}} \mathbf{H}_{\alpha\alpha}^R e^{X_{\alpha 1}(a_{\alpha}^+ - 1)} | 0 \rangle + \sum_{\beta \neq \alpha} \langle \tilde{\phi}_{\beta} | \mathbf{H}_{\alpha\beta}^R | \psi_{\beta} \rangle \quad (31)$$

$$H_{total}^P := \langle 0 | e^{a_{\alpha}} \mathbf{H}_{\alpha\alpha}^P e^{X_{\alpha 0}(a_{\alpha}^+ - 1)} | 0 \rangle + \sum_{\beta \neq \alpha} \langle \tilde{\phi}_{\beta} | \mathbf{H}_{\alpha\beta}^P | \psi_{\beta} \rangle \quad (32)$$

$$f_{total}^{R(P)} := f_{\alpha\alpha}^{R(P)} + \sum_{\beta \neq \alpha} f_{\alpha\beta}^{R(P)} \quad (33)$$

$$H_{total}^R := \langle 0 | e^{a_{\alpha}} a_{\alpha}^+ \mathbf{H}_{\alpha\alpha}^R e^{X_{\alpha 1}(a_{\alpha}^+ - 1)} | 0 \rangle + \sum_{\beta \neq \alpha} \langle \tilde{\phi}_{\beta} | \mathbf{H}_{\alpha\beta}^R | \psi_{\beta} \rangle X_{\alpha 1} \quad (34)$$

$$H_{total}^P := \langle 0 | e^{a_{\alpha}} a_{\alpha}^+ \mathbf{H}_{\alpha\alpha}^P e^{X_{\alpha 0}(a_{\alpha}^+ - 1)} | 0 \rangle + \sum_{\beta \neq \alpha} \langle \tilde{\phi}_{\beta} | \mathbf{H}_{\alpha\beta}^P | \psi_{\beta} \rangle X_{\alpha 1} \quad (35)$$

where  $\langle \tilde{\phi}_{\beta} |$  is defined by Equation 19. We can interpret Equation 27 as a master equation of probabilities of active operator states related to reactions Equation 1–4. However its change of the rate is based on the average number of proteins (Equations 31–32) unlike usual master equations which deal with the specific state. Equation 28 is just probability conservation. The first term of Equation 29 and Equation 30 is related to protein synthesis and degradation (Equations 5–7), and the rest are the effects of operator fluctuations.

Equations 27–30 are the general equations for parameters; we can use these equations for any number of gene systems and for any kind of regulatory patterns by assigning the binding and unbinding probability rate.

**Moments equation.** The mean-field approximation approach should inherently provide information on moments (Equation 12), so it is natural to construct moment equations from Equation 11 [12]. The  $k$ -th moment of the protein number is obtained by the following equations. The 0-th moment of protein number is interpreted as the probability that gene is on or off. We also derived the moments equations from Equation 18

$$\left( \langle 0 | e^{a_{\alpha}} (a_{\alpha}^+ a_{\alpha})^k | 0 \rangle \otimes \prod_{\beta \neq \alpha} \langle \tilde{\phi}_{\beta} | \right) \left( \frac{\partial}{\partial t} |\Psi\rangle - \Omega |\Psi\rangle \right) = 0 \quad (36)$$

$$\left( \langle 0 | e^{a_{\alpha}} (a_{\alpha}^+ a_{\alpha})^k | 0 \rangle \otimes \prod_{\alpha \neq \beta} \langle \tilde{\phi}_{\beta} | \right) \left( \frac{\partial}{\partial t} |\Psi\rangle - \Omega |\Psi\rangle \right) = 0 \quad (37)$$

In principle, once we know all the moments, we can construct a probability distribution; but in many cases, we cannot get all the moments. However, Equations 36 and 37 have a good structure if there is no self-interaction, since all the moments can be computed by some lower moments recursively. If we deal with  $q$ -th multimer proteins, we can get all the moments higher than  $q$  from 0-th, . . . ,  $q$ -th moments. In the toggle switch and the dimer protein cases, for example, we have 12 independent equations ( $3 \times 2 \times 2$ : ( $k = 0, 1, 2$ ), ( $\alpha = A, B$ )) for each Equation 36 and Equation 37) and 12 variables ( $C_{\alpha 1}, C_{\alpha 0}, \langle n_{\alpha 1} \rangle, \langle n_{\alpha 0} \rangle, \langle n_{\alpha 1}^2 \rangle, \langle n_{\alpha 0}^2 \rangle, \alpha = A, B$ ); we show these explicitly in Equations 45–50, where we used two constraints of probability conservation, so there are only ten equations. We can say these equations are closed. However, if we include self-interaction, the equations are not closed, so they cannot be solved exactly [12].

**Relation between the ansatz equation and the moments equation.** Equations 27–30 can be derived from Moment Equations 36 and 37 without using the variational principle. We can start from moment equations and then assume a specific probability distribution based on a physical argument, which gives some specific relations between moments. For example, the Poisson distribution has only one parameter, so we may calculate all other moments from the first moment, the mean. Moment equations with a Poisson ansatz give us the same equations as the variational approach in Equations 27–30.

Moment equations are more exact than the variational approach, but the approach cannot be used to obtain exact solutions for the system having self-interaction, in which equations are not closed. Even in the closed system, an ansatz reduces the degrees of freedom significantly and makes the problem easier to handle. Mathematically, using an ansatz is equivalent to giving specific relations between moments. We may, therefore, not need to take care of higher moments if an infinite number of higher moments is automatically given by assuming a specific ansatz. In practice, ansatz might be useful. Then the issue would be how faithful the ansatz we choose is. In this paper we used both the moment equation and the Poisson ansatz. Notice that Equation 11 is merely a birth–death process without the last term. In the limit that the last term is small enough, the so-called adiabatic limit (faster protein number fluctuations compared with the DNA state alterations), we expect the solution will be close to the Poisson distribution.

## Interpretation of the Solutions

The final output we get from these equations is basically moments. From these moments we need to construct the total probability. There are several important features to be pointed out. We start with the single gene case.

First, notice that the total probability does not have the structure of  $C_1 P_1 + C_0 P_0$ . We started with a two-component column vector and to extract the physical observables we needed to do the inner product with a two-component row state vector. (We never added the spin up and down component directly in quantum mechanics.) The total probability should therefore not follow the steps of constructing  $P_1$  and  $P_0$  first and then weighing by  $C_1$  and  $C_0$ . The correct procedure is the following. With the moments, the solutions of equations, we construct new moments:

$$\begin{aligned}
1 &= C_1 + C_0 \\
\langle n \rangle &= C_1 \langle n \rangle_{on} + C_0 \langle n \rangle_{off} \\
\langle n^k \rangle &= C_1 \langle n^k \rangle_{on} + C_0 \langle n^k \rangle_{off} \quad k = 2, 3, \dots
\end{aligned}$$

In principle, we can get arbitrary order of moments and construct the corresponding probability if the equations are closed. In practice, however, we may choose one of two probability distributions: Poisson or Gaussian distributions.

Second, the probability obtained above corresponds to one limit point or basin of attraction. One solution of the equations determines one of the limit points and also gives the variation around the basin of attraction, so it is intrinsic. If the system allows multistability, then there are several probability distributions localized at each basin of attraction, but with different variations. Thus, the total probability is the weighted sum of all these probability distributions. The weighting factors ( $w^a$ ,  $w^b$ ) are the size of the basin, which is nothing but the relative size of the set of initial values ending up with a specific basin of attraction.

$$P(n) = w^a P^a(n) + w^b P^b(n) \quad w^a + w^b = 1. \quad (38)$$

Notice that the steady state solution is not enough to describe the total probability. It does not say anything about the volume of the basin, it only tells us the limit point. So the effort to derive an effective potential energy from the steady state solution on general grounds needs to take into account the volume of the basin of attraction. One simple exception is the symmetric toggle switch, where the weighting factors are simply (0.5, 0.5) by symmetry.

Third, the total probability of many genes is simply the product of each gene based on our basic assumption, the mean field approximation. For example, the probability of a toggle switch can be written as

$$P(n_A, n_B) = w^a P^a(n_A) P^a(n_B) + w^b P^b(n_A) P^b(n_B) \quad (39)$$

where  $a$  and  $b$  denote each limit point, and  $w^a$  and  $w^b$  are the weighting factors. Even though it is simply multiplication, the interactions between them are already taken into account from the coupled equations.

Finally, once we have the total probability, we can construct the potential energy (or potential energy landscape) by the relationship with the steady state probability:

$$U(\vec{n}) = -\ln P(\vec{n}). \quad (40)$$

This is the reverse order of the usual statistical mechanics of first obtaining the potential energy function, exponentially Boltzman weighting it, and then studying the partition function or probability of the associated system. Here we look for the inherent potential energy function from the steady state probability. In the gene-network system, every chemical parameter, such as the protein production/decay rates and binding/unbinding rates, will contribute to the fluctuation of the system. All these effects are encoded in the total probability distribution, and, consequently, in the underlying potential energy landscape.

## Results

We looked at an important example of two genes interacting with each other. The interactions are through the proteins synthesized by the genes, which act back to regulate the gene switch. The bacterial lambda phage is a good biological example of a toggle switch. The two lysogenic

and lysosic genes are both stable and robust. It has been a long-standing problem to explain why the lambda phage is so stable [11,12,16,34,40]. We addressed this issue in the presence of the intrinsic statistical fluctuations of the finite number of the proteins in the cell by exploring the underlying potential energy landscape. First we solved the master equation to obtain both the time-dependent and steady state probability distributions of the protein concentrations of the corresponding genes. Then we inferred the underlying potential energy landscape from the steady state probability distribution of the protein concentrations. We then considered the symmetric toggle switch.

All applications to specific network systems start with Equations 27–30. First, we chose the number of genes and designed the interaction type (network topology) and protein types (multimers). Second, we assigned the strength of the parameters. Then we solved this coupled ordinary differential equation system numerically with certain initial conditions. We considered a toggle switch [39] of two genes, as shown in Figure 1, which has wide application in molecular biology, such as the bacteriophage  $\lambda$  problem [34]. Let us start with the toggle switch case.

### Symmetric Toggle Switch

For the symmetric switch, we first solved the equations of motion determining the amplitude, the mean, and the higher order moments of the probability distribution of the protein concentrations of the corresponding genes. These are given below.

### Poisson and Moment Equation Solutions of Master Equations

We solved the master equation with two methods. One is the Poisson ansatz, mentioned above, by assuming the inherent Poisson distribution, and the other is the exact method, using the moment equation. For the inherent Poisson distribution, we can write down the equations of motion for the amplitude and mean.

#### Poisson ansatz.

$$\frac{dC_{A1}}{dt} = -\frac{h_A}{2} C_{A1} \{C_{B1} X_{B1}^2 + (1 - C_{B1}) X_{B0}^2\} + f_A (1 - C_{A1}) \quad (41)$$

$$\frac{dX_{A1}}{dt} = -f_A \frac{1 - C_{A1}}{C_{A1}} (X_{A1} - X_{A0}) + g_{A1} - k_A X_{A1} \quad (42)$$

$$\frac{dX_{A0}}{dt} = f_A (X_{A1} - X_{A0}) + g_{A0} - k_A X_{A0} - \frac{X_{A1} - X_{A0}}{1 - C_{A1}} C_{A1} \quad (43)$$

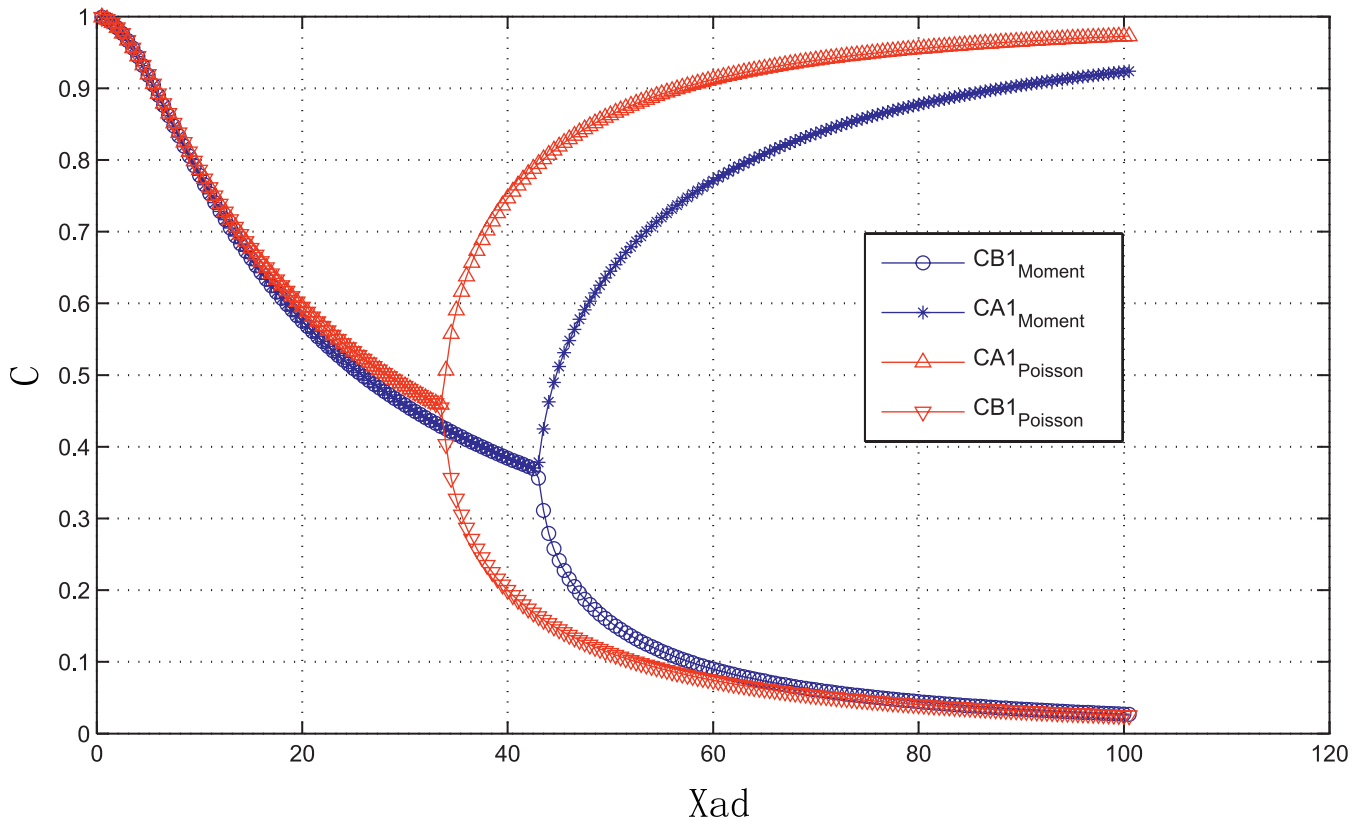
$$\text{three - more - equations}(A \leftrightarrow B) \quad (44)$$

where we eliminated two variables by the probability conservation ( $C_{x1} + C_{x0} = 1$ ), and recollected terms.

For the exact solution with moment equations, we also wrote down the equations of motion of the moment of protein concentration of the corresponding genes.

#### The corresponding moment equations.

$$\begin{aligned}
\frac{dC_{A1}}{dt} &= -\frac{h_A}{2} C_{A1} \{C_{B1} (\langle n_{B1}^2 \rangle - \langle n_{B1} \rangle) + (1 - C_{B1}) (\langle n_{B0}^2 \rangle - \langle n_{B0} \rangle)\} \\
&\quad + f_A (1 - C_{A1})
\end{aligned} \quad (45)$$



**Figure 2.** Probability  $C$  That Genes Are in the Active State as a Function of  $X_{ad} = (g_1 + g_0)/2k_A$  for a Symmetric Switch Showing the Bifurcation Exact moment equation solutions are compared with Poisson ansatz solutions  $0 < X_{ad} (= g_1/2) < 100$ ,  $k_A = k_B = 1$ ,  $f_A = f_B = 0.5$ ,  $h_A = h_B = f_A/500$ , and  $g_{A0} = g_{B0} = 0$ . doi:10.1371/journal.pcbi.0030060.g002

$$\frac{d\langle n_{A1} \rangle}{dt} = -f_A \frac{1 - C_{A1}}{C_{A1}} (\langle n_{A1} \rangle - \langle n_{A0} \rangle) + g_{A1} - k_A \langle n_{A1} \rangle \quad (46)$$

$$\frac{d\langle n_{A0} \rangle}{dt} = f_A (\langle n_{A1} \rangle - \langle n_{A0} \rangle) + g_{A0} - k_A \langle n_{A0} \rangle - \frac{\langle n_{A1} \rangle - \langle n_{A0} \rangle}{1 - C_{A1}} C_{A1} \quad (47)$$

$$\begin{aligned} \frac{dC_{A1}}{dt} \langle n_{A1}^2 \rangle + C_{A1} \frac{d\langle n_{A1}^2 \rangle}{dt} &= C_{A1} g_{A1} (2\langle n_{A1} \rangle + 1) + k_A C_{A1} (-2\langle n_{A1}^2 \rangle \\ &+ \langle n_{A1} \rangle) + f_A (1 - C_{A1}) \langle n_{A0}^2 \rangle \\ &- C_{A1} \frac{h_A}{2} \langle n_{A1}^2 \rangle (C_{B1} (\langle n_{B1}^2 \rangle - \langle n_{B1} \rangle) \\ &+ (1 - C_{B1}) (\langle n_{B0}^2 \rangle - \langle n_{B0} \rangle)) \end{aligned} \quad (48)$$

$$\begin{aligned} -\frac{dC_{A1}}{dt} \langle n_{A0}^2 \rangle + (1 - C_{A1}) \frac{d\langle n_{A0}^2 \rangle}{dt} &= (1 - C_{A1}) g_{A0} (2\langle n_{A0} \rangle + 1) \\ &+ k_A (1 - C_{A1}) (-2\langle n_{A0}^2 \rangle \\ &+ \langle n_{A0} \rangle) - f_A (1 - C_{A1}) \langle n_{A1}^2 \rangle \\ &+ C_{A1} \frac{h_A}{2} \langle n_{A1}^2 \rangle (C_{B1} (\langle n_{B1}^2 \rangle \\ &- \langle n_{B1} \rangle) + (1 - C_{B1}) (\langle n_{B0}^2 \rangle \\ &- \langle n_{B0} \rangle)) \end{aligned}$$

$$\text{five - more - equations } (A \leftrightarrow B) \quad (50)$$

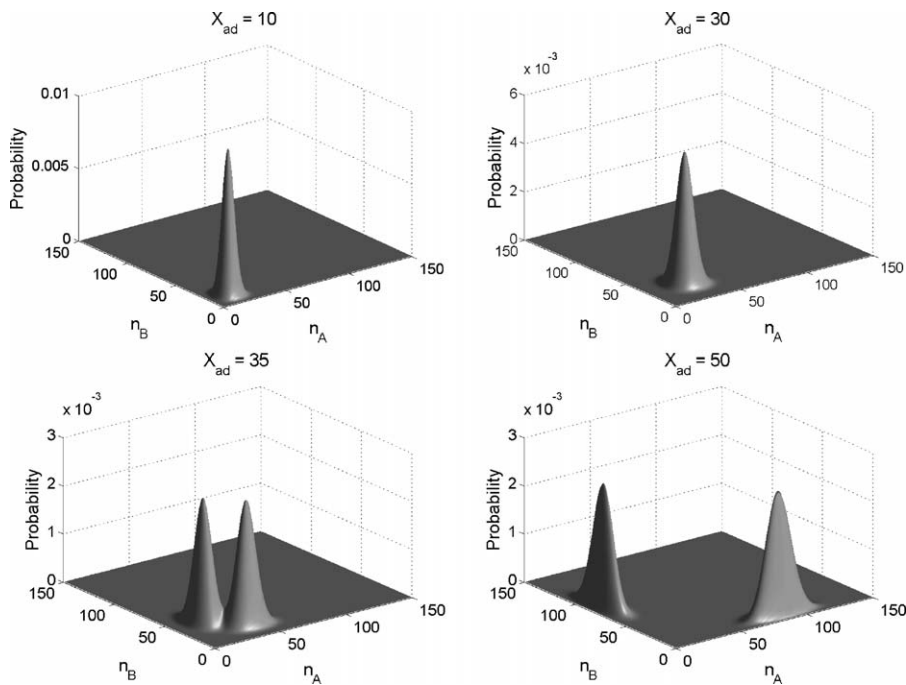
### Monostability versus Bistability for Symmetric Toggle Switch

By giving some initial conditions, and taking the long time limit, we obtained the steady state solution. We fixed all parameters except the protein synthesis rate  $g_{A1} (= g_{B1})$ . We looked at the probability of genes that were in the active state versus the relative importance of synthesis rate versus degradation rate. By increasing the synthesis rate,  $g_{A1}$ , we could observe the bifurcation from the monostable state to the bistable state after passing a certain critical point. Figure 2 shows the result of taking the long time limit of the equations of motion. The two curves (with subscript Moment) are from moment equations, and the others are from the Poisson ansatz. This is consistent with the results directly from time-independent equations (Figure 3A in [12]). We used the parameter  $X_{ad} = \bar{g}/k$  as the horizontal axis variable. It is simply  $g_1/2$  in our choice.

In the parameter range in which the bistability occurs, we found two limit points (named  $a$  and  $b$ ) in the numerical analysis. Now from the solution of the equations we constructed the probability of protein numbers or concentrations (all illustrations in this paper were based on the Poisson ansatz for simplicity, but it can be easily done with the moment equations and qualitative features will not be changed),

$$P(n_A, n_B) = w^a P^a(n_A) P^a(n_B) + w^b P^b(n_A) P^b(n_B) \quad (51)$$

For the symmetric toggle switch case, the weight factor was



**Figure 3.** Steady State Probability of Symmetric Toggle Switch (Long Time Limit) as a Function of the Number of Protein A and the Number of Protein B for Different  $X_{ad} = g_{A1}/2k_A$

The other parameters are the same as Figure 2.  
doi:10.1371/journal.pcbi.0030060.g003

simply (0.5, 0.5) due to symmetry. The change of the probability distribution shape in terms of the adiabatic parameter of the relative importance of the protein synthesis rate compared with the degradation rate is shown in Figure 3. These figures show the monostability to bistability of the symmetric toggle switch. For a large enough protein synthesis rate relative to degradation rate, bistability emerges.

#### Potential Energy Landscape: Monostability to Bistability

As we discussed, the steady-state distribution function  $P\vec{n}$  for the state variable  $\vec{n}$  can be expressed to be exponential in a function  $U\vec{n}$ :

$$P(\vec{n}) = \exp\{-U(\vec{n})\}, \quad (52)$$

where  $P\vec{n}$  is already normalized. From the steady state distribution function, we can therefore identify  $U$  as the generalized potential energy function of the network system. In this way, we map out the potential energy landscape. Figure 4 shows the potential energy landscape corresponding to Figure 3.

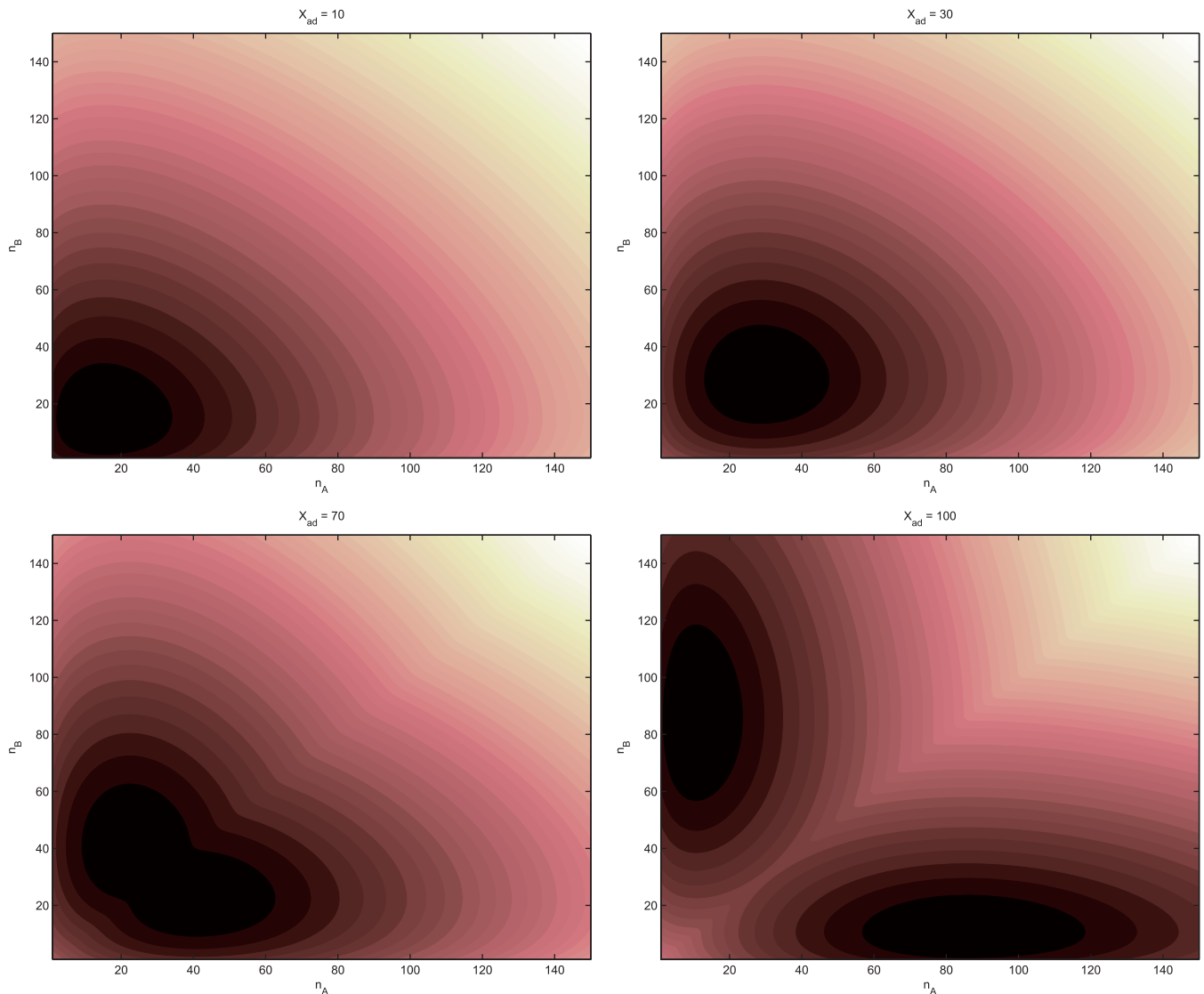
We can see that when the protein synthesis rate is small relative to degradation rate, only a single basin of attraction exists for the underlying potential energy landscape. For large enough protein synthesis rate relative to degradation rate, two basins of attraction emerge. Once we have the potential energy landscape, we can discuss the global stability of the gene regulatory networks. The time scale of the transition between the two stable minimum basins of attraction can be estimated by  $\tau \sim \tau_0 \exp[U^\ddagger - U_{min}]$  [41]. Here,  $\tau_0$  is the pre-factor and  $\tau$  is the time scale of transition from one basin of attraction to the other.  $U^\ddagger$  is the potential energy at the saddle point between the two stable basins of attraction.  $U_{min}$  is the potential energy at one of the basins of

attraction. Thus  $U^\ddagger - U_{min}$  represents the potential energy barrier height between two stable basins of attraction. In Figure 5 we can see that as the synthesis rate and unbinding rate of protein to DNA increase relative to the degradation rate, the potential barrier height between the two basins of attraction increases. The time scale of the transition from one basin of attraction to the other exponentially increases with the barrier height.

Figure 5 shows the phase diagram of the parameter ranges for the monostable basin and two bistable basins of attraction. We can see that when the synthesis rate and unbinding rate of protein to DNA are low relative to the degradation rate, the potential energy landscape prefers one stable basin of attraction. As the synthesis rate and unbinding rate of protein to DNA increases relative to the degradation rate, the potential energy landscape gradually develops the two stable basins of attraction from the monostable one. There is a transition from monostable to bistable basins of attraction of the underlying potential energy landscape at certain parameters.

This illustrates how biological robustness is realized for the toggle switch. As the protein synthesis rate and unbinding rate of protein to DNA increase relative to the degradation rate, more proteins are synthesized. These proteins are strong repressors. This leads to smaller fluctuations. Furthermore, the associated barrier height between the two basins of attraction becomes large, and the two basins of attraction become more stable since it is harder to go from one well to another. So, small fluctuations and large barrier heights both serve as the source for the robustness and stability of the gene toggle switch. In other words, it is more unlikely for the system to change from one basin of attraction to the other. Therefore, the system becomes robust. The robustness issue is





**Figure 4.** Potential Energy of Symmetric Toggle Switch as a Function of the Number of Protein A and the Number of Protein B for Different  $X_{ad} = g_{A1}/2k_A$

The other parameters are the same as Figure 2.  
doi:10.1371/journal.pcbi.0030060.g004

not yet well-understood for cellular networks in general. Here we explored the robustness of the switches against the intrinsic statistical fluctuations coming from the finite number of protein and DNA molecules. This is clearly very important and has potential applications to the robustness problem of lambda phage in bacteria.

#### Time Evolution of the Underlying Potential Energy Landscape

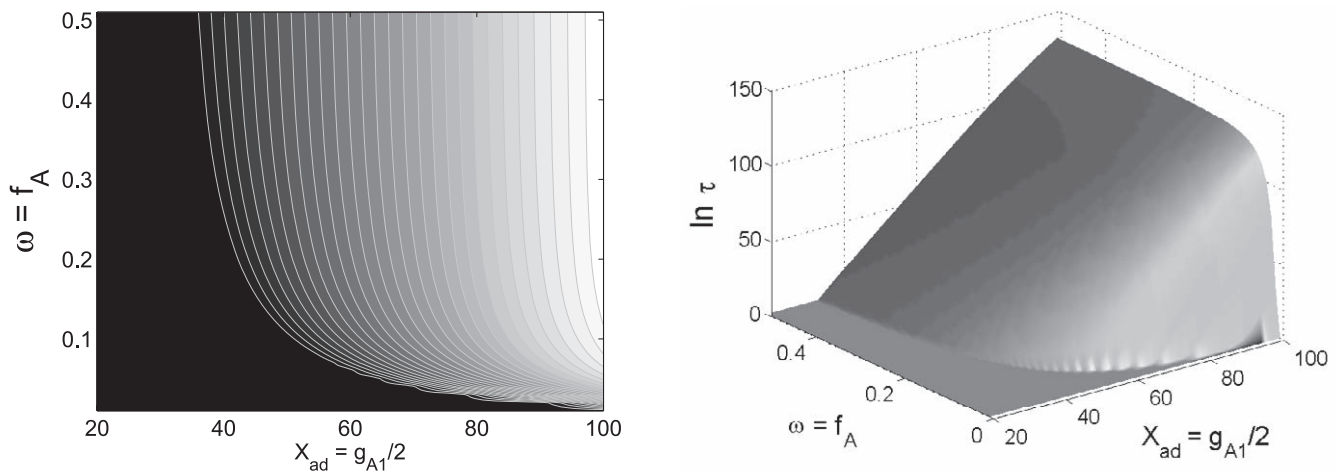
We also studied the time evolution of the probability and the potential energy landscape with dynamic equations. We chose the specific parameters and initial conditions to illustrate the idea. The results are shown in Figures 6 and 7.

In Figures 6 and 7, we see the evolution in time of the probability and the underlying potential energy landscape from the flat land at the beginning to the full development of two basins of attraction at the steady state. This is the first illustration of the dynamical evolution or formation of

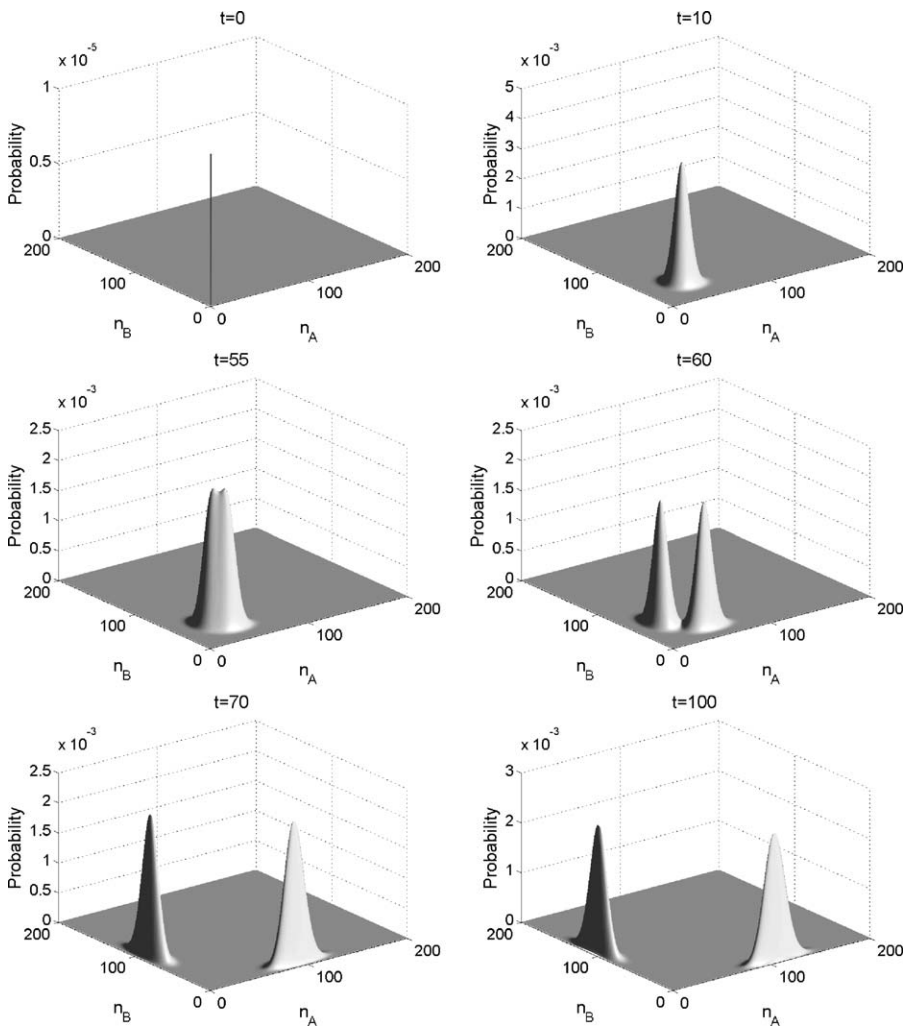
development of the potential energy landscape of a toggle switch.

#### Discussion

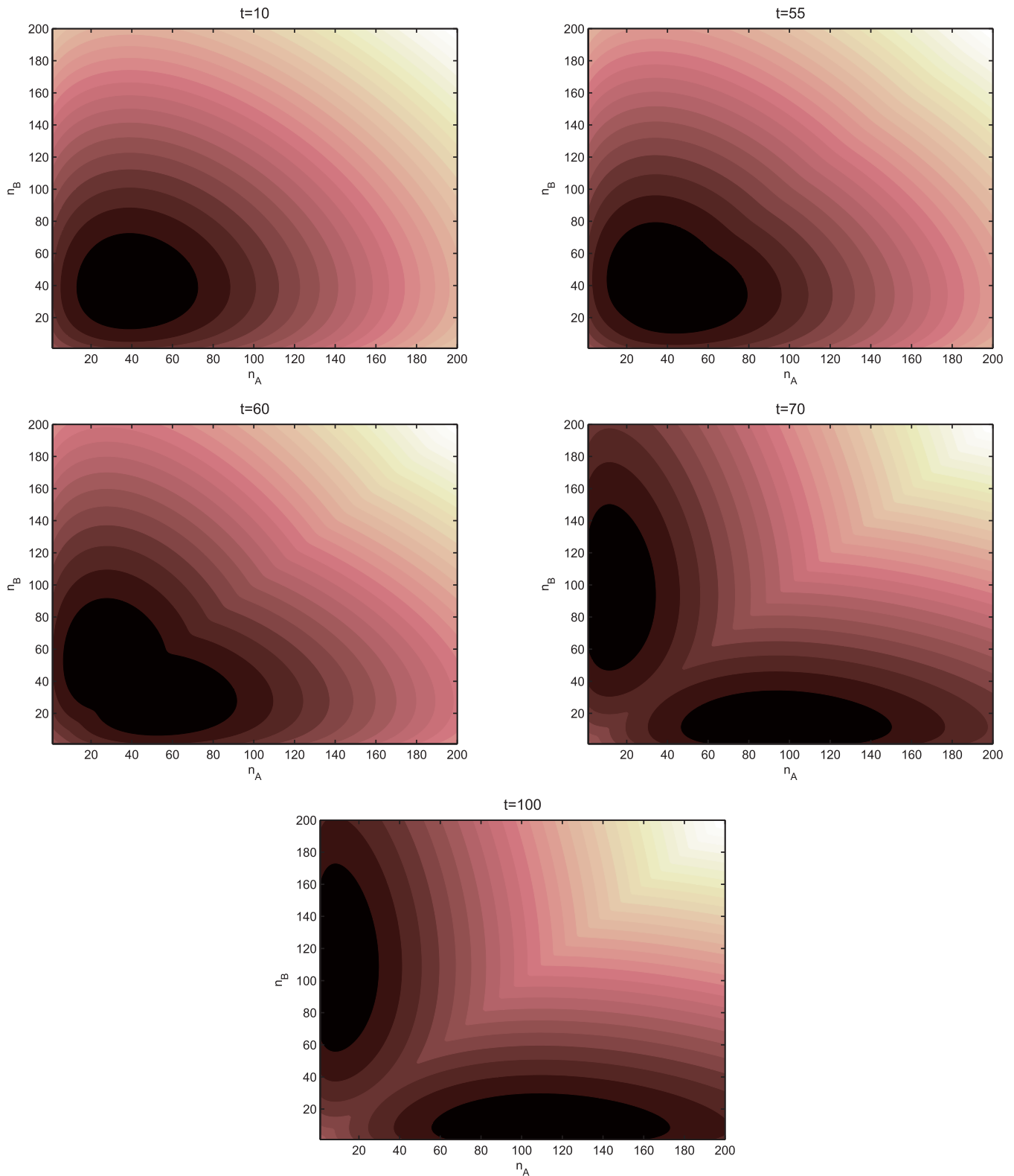
Finding the multidimensional potential energy landscape is the key to addressing important global issues such as the robustness of cellular networks. We have uncovered the underlying potential energy landscape of a simple gene network: toggle switch. We found that as the protein synthesis rate and the unbinding of protein to DNA rate relative to degradation change from small to large, the underlying potential energy landscape changes from having monostable to bistable basins of attraction. These basins correspond to stable, biologically functional states. The potential barrier between the two basins determines the time scale of conversion from one to the other. We found that as the protein synthesis rate and unbinding of protein with DNA



**Figure 5.** The Time Scale of the Transition between the Two Stable Minimum Basins of Attractions as a Function of  $\omega = f_A/k_A$  and  $X_{ad} = g_{A1}/2k_A$ . The other parameters are the same as Figure 2. doi:10.1371/journal.pcbi.0030060.g005



**Figure 6.** The Time Evolution of the Probability of Symmetric Toggle Switch as a Function of the Numbers of Protein A and the Numbers of Protein B  $X_{ad} = 60$ , and the other parameters are the same as Figure 2.  $t = 0, 10, 55, 60, 70, 100$ . doi:10.1371/journal.pcbi.0030060.g006



**Figure 7.** The Time Evolution of the Potential of Symmetric Toggle Switch as a Function of the Numbers of Protein A and the Numbers of Protein B  $X_{ad} = 60$ , and the other parameters are the same as Figure 2.  $t = 10, 55, 60, 70, 100$ .  
doi:10.1371/journal.pcbi.0030060.g007

rate relative to degradation became greater, the potential energy barrier became greater and the statistical fluctuations were effectively more severely suppressed. This leads to the robustness of the biological basins of the gene switches.

In principle, our approach can be generalized to more realistic networks involving multiple genes as well as additional levels of regulations. This could be realized by averaging the interactions among genes in the corresponding

master equations. It effectively reduces the dimensionality of the problem from exponential to polynomial number of degrees of freedom. It is worthwhile to note the limitation of this approach. When the interactions among genes are very strong, our approach is less effective.

Recently, synthetic biology became an important part of systems biology [42–45]. There has been significant progress in this field. However, there still seems to be a lack of general principles and algorithms guiding the design and construction of synthetic gene networks. The robustness condition (see Figure 5) found in this study would help us to identify the parameter and connectivity region to reach global robustness and function of the network. The optimal network design will be based on that. Furthermore, we can vary the parameters and connections to design different distinct features while maintaining the stability of the network.

The adaptive landscape idea was first introduced into biology by S. Wright in the 1930s [46–49]. Landscape construction for one dimension is rather straightforward. However, even the two-dimensional case becomes nontrivial. The recent efforts to understand global systems biology need the concept of landscape. Progress was made towards this from the dynamic system point of view, where the nontrivial nature

of low-dimensional systems was illustrated [20,36,50]. There are still conceptual and methodological issues remaining for high dimensional systems. The stochastic method introduced here may pave the road towards solving this problem.

This model can be modified to include more biochemical reactions. To investigate the role of mRNA, we can consider the transcription and translation process separately. To focus on the statistical fluctuations of genes turning on and off, it is possible to generalize the formalism to compute the statistical fluctuations quantitatively. We also can take into account the spatial variation of the state variables, such as the number of proteins.

## Acknowledgments

JW would like to thank Professor Peter G. Wolynes, Professor Ping Ao, and Dr. Xiaomei Zhu for helpful discussions.

**Author contributions.** JW conceived and designed the experiments. KYK performed the experiments. KYK and JW analyzed the data, contributed reagents/materials/analysis tools, and wrote the paper.

**Funding.** This work was supported by grants from the US National Science Foundation Career Award, the Petroleum Research Fund, and the National Science Foundation of China.

**Competing interests.** The authors have declared that no competing interests exist.

## References

- Davidson EH (2006) The regulatory genome: Gene regulatory networks in development and evolution. New York: Academic Press. 304 p.
- Huang CY, Ferrell JE Jr (1996) Ultrasensitivity in the mitogen-activated protein kinase cascade. *Proc Natl Acad Sci U S A* 93: 10078–10082.
- Kholodenko BN (2000) Negative feedback and ultrasensitivity can bring about oscillations in the mitogen-activated protein kinase cascades. *Eur J Biochem* 267: 1583–1593.
- Ideker T, Thorsson V, Ranish JA, Christmas R, Buhler J, et al. (2001) Integrated genomic and proteomic analyses of a systematically perturbed metabolic network. *Science* 292: 929–933.
- Jeong H, Tombor B, Albert R, Oltvai ZN, Barabasi AL (2000) The large-scale organization of metabolic networks. *Nature* 407: 651–654.
- Barabasi AL, Bonabeau E (2003) Scale-free networks. *Sci Am* 288: 60–69.
- Maslov S, Sneppen K (2002) Specificity and stability in topology of protein networks. *Science* 296: 910–913.
- Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, et al. (2002) Network motifs: Simple building blocks of complex networks. *Science* 298: 824–827.
- Tyson JJ, Chen K, Novak B (2001) Network dynamics and cell physiology. *Nat Rev Mol Cell Biol* 2: 908–916.
- Bialek W (2003) Stability and noise in biochemical switches. *Adv Neural Infor Process* 13: 103–109.
- Sasai M, Wolynes PG (2003) Stochastic gene expression as a many-body problem. *Proc Natl Acad Sci U S A* 100: 2374–2379.
- Walczak AM, Sasai M, Wolynes PG (2005) Self consistent proteomic field theory of stochastic gene switches. *Biophys J* 88: 828–850.
- Hornos JE, Schultz D, Innocentini GC, Wang J, Walczak AM, et al. (2005) Self-regulating gene: An exact solution. *Phys Rev E Stat Nonlin Soft Matter Phys* 72: 051907.
- Li F, Long T, Lu Y, Ouyang Q, Tang C (2004) The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci U S A* 101: 4781–4786.
- Ao P (2004) Potential in stochastic differential equations: Novel construction. *J Phys A Math Gen* 37: L25–L30.
- Zhu XM, Yin L, Hood L, Ao P (2004) Calculating biological behaviors of epigenetic states in the phage lambda life cycle. *Funct Integr Genomics* 4: 188–195.
- Qian H, Bear DA (2005) Thermodynamics of stoichiometric biochemical networks far from equilibrium. *Biophys Chem* 114: 213–220.
- Qian H, Reluga TC (2005) Nonequilibrium thermodynamics and nonlinear kinetics in a cellular signaling switch. *Phys Rev Lett* 94: 028101.
- Wang J, Huang B, Xia XF, Sun ZR (2006) Funneled landscape leads to robustness of cellular networks: MAPK signal transduction. *Biphs J Lett* 91: L54–L56.
- Wang J, Han B (2007) Quantifying robustness of cell cycle network: Funneled energy landscape perspectives. *Biophys J*. In press.
- Austin RH, Beeson K, Eisenstein L, Frauenfelder H, Gunsalus I, et al. (1975) Ligand binding to myoglobin. *Biochem* 14: 5355–5373.
- Frauenfelder H, Sligar SG, Wolynes PG (1991) The energy landscapes and motions of proteins. *Science* 254: 1598–1603.
- McAdams HH, Arkin A (1997) Stochastic mechanisms in gene expression. *Proc Natl Acad Sci U S A* 94: 814–819.
- Elowitz MB, Leibler S (2000) A synthetic oscillatory network of transcriptional regulators. *Nature* 403: 335–338.
- Swain PS, Elowitz MB, Siggia ED (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci U S A* 99: 12795–12800.
- Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci U S A* 98: 8614–8619.
- Vilar JMG, Guet CC, Leibler S (2003) Modeling network dynamics: The lac operon, a case study. *J Cell Biol* 161: 471–476.
- Paulsson J (2004) Summing up the noise in gene networks. *Nature* 427: 415–418.
- Hasty J, Pradines J, Dolnik M, Collins JJ (2000) Noise-based switches and amplifiers for gene expression. *Proc Natl Acad Sci U S A* 97: 2075–2080.
- Hasty J, Isaacs F, Dolnik M, McMillen D, Collins JJ (2001) Designer gene networks: Towards fundamental cellular control. *Chaos* 11: 207–220.
- Gardiner CW (1985) Handbook of stochastic methods for physics, chemistry and the natural sciences. Berlin: Springer-Verlag. 475p.
- van Kampen NG (1992) Stochastic processes in chemistry and physics. Amsterdam: North-Holland. 419 p.
- Gillespie DT (1977) Exact stochastic simulation of coupled chemical reactions. *J Phys Chem* 81: 2340–2361.
- Arkin A, Ross J, McAdams HH (1998) Stochastic kinetic analysis of developmental pathway bifurcation in phage  $\lambda$ -infected *Escherichia coli* cells. *Genetics* 149: 1633–1649.
- Kepler TB, Elston TC (2001) Stochasticity in transcriptional regulation: Origins, consequences, and mathematical representations. *Biophys J* 81: 3116–3136.
- Kim K, Lepzelter D, Wang J (2007) Single molecule dynamics, statistical fluctuations, amplitude and period of oscillations of a gene regulatory network: Repressilator. *J Chem Phys*. 126: 1–11.
- Mattis DC, Glasser ML (1998) The uses of quantum field theory in diffusion-limited reactions. *Rev Mod Phys* 70: 979–1001.
- Eyink GL (1996) Action principle in nonequilibrium statistical dynamics. *Phys Rev E* 54: 3419–3435.
- Gardner TS, Cantor CR, Collins JJ (2000) Construction of a genetic toggle switch in *Escherichia coli*. *Nature* 403: 339–342.
- Aurell E, Sneppen K (2002) Epigenetics as a first exit problem. *Phys Rev Lett* 88: 048101.
- Walczak AM, Onuchic JN, Wolynes PG (2005) Absolute rate theories of epigenetic stability. *Proc Natl Acad Sci U S A* 102: 18926–18931.
- You L, Cox RS III, Weiss R, Arnold FH (2004). Programmed population control by cell–cell communication and regulated killing. *Nature* 428: 868–871.
- Guido NJ, Wang X, Adalsteinsson D, McMillen D, Hasty J, et al. (2006) A bottom-up approach to gene regulation. *Nature* 439: 856–860.
- Guet CC, Elowitz MB, Hsing W, Leibler S (2002) Combinatorial synthesis of genetic networks. *Science* 296: 1466–1470.
- Kobayashi M, Iaccarino C, Saiardi A, Heidt V, Bozzi Y, et al. (2004)

- Simultaneous absence of dopamine D1 and D2 receptor-mediated signaling is lethal in mice. *Proc Natl Acad Sci U S A* 101: 11465–11470.
46. Fisher RA (1930) *The genetical theory of natural selection*. Oxford: Clarendon. 251 p.
  47. Wright S (1932) The roles of mutation, inbreeding, crossbreeding and selection in evolution. In: *Proceedings of the Sixth International Congress on Genetics*. 1: 356–366.
  48. Delbruck M (1949) Discussion. In: *Unites Biologiques Douces de Continuite Genetique Colloques Internationaux du Centre National de la Recherche Scientifique*. Paris: CNRS.
  49. Waddington CH (1957) *Strategy of the gene*. London: Allen and Unwin. 290 p.
  50. Ao P (2005) Laws in Darwinian evolutionary theory. *Phys Life Rev* 2: 117–156.