

## Stochastic model of coliphage lambda regulatory network

Siyuan Wang,<sup>1,2</sup> Yuping Zhang,<sup>1,3</sup> and Qi Ouyang<sup>1,2,\*</sup>

<sup>1</sup>*Center for Theoretical Biology, Peking University, Beijing, 100871, China*

<sup>2</sup>*Department of Physics, Peking University, Beijing, 100871, China*

<sup>3</sup>*Department of Mathematics, Peking University, Beijing, 100871, China*

(Received 22 April 2005; revised manuscript received 28 November 2005; published 20 April 2006)

The dynamic properties of the regulatory network governing the choice between lytic and lysogenic growths of coliphage lambda is studied using a Markov chain stochastic model. Our computer simulation confirms the finding by Li *et al.* [Proc. Natl. Acad. Sci. USA **101**, 4781 (2004)] on the dynamics of budding yeast: that the biological stationary states are global attractors; the biological pathways of lytic and lysogenic growths are attracting trajectories; and the network functions are robustly designed against structural perturbations. In addition, our model shows the stochastic switch from lysogen to lytic growth, which has been observed in experiments. A definition of pseudoenergy is introduced in the network dynamics to reveal a transitionlike behavior in the system.

DOI: [10.1103/PhysRevE.73.041922](https://doi.org/10.1103/PhysRevE.73.041922)

PACS number(s): 87.17.Aa, 87.17.-d, 87.10.+e, 87.16.Yc

### I. INTRODUCTION

Current biological network studies cover from macroscopic food webs, in which the nodes are species and the edges represent predator-prey relationships between them, to microscopic metabolism networks, in which the nodes are proteins and substrates, and the edges represent chemical reactions [1]. Studies have shown that biological networks have many special topological characteristics that are not seen in random networks of the same size (i.e., with the same numbers of nodes and edges), such as stability, high clustering coefficient, and tolerance for errors and attacks [1]. Recently, Li *et al.* [2] reported that the biological network of the budding yeast cell-cycle process possesses special dynamic properties: The biological stationary state is a global attractor; the biological pathway is an attracting trajectory; and the network functions are robustly designed against structural perturbations. In this report we study the dynamics of another biological network, the growth modes of bacteriophage lambda, in order to verify the universality of the dynamical properties mentioned above.

Bacteriophage lambda is a well-studied temperate phage, and has long been used as a model system among biologists. Different dynamic models have been established to analyze the stability of its genetic switch [3–15], which help us to quantitatively understand the dynamic process of bacteriophage lambda's growth. However, most previous studies omit fluctuations in the system, which might play an important role in a living cell. From the biological point of view, one realizes that there are relatively few copies of each protein and RNA in a single cell, and each molecule evolves stochastically. To characterize such a system, it is more reasonable to apply a stochastic model rather than a deterministic one. In order to give a microscopic statistical understanding of the system, the recent trend in lambda phage simulation is to involve stochastic analysis [12–15]. Three kinds of ap-

proaches have been documented: Ref. [15] offers fluctuation to the degradation of protein CI in the standard ordinary differential equation (ODE) model in order to reveal the developmental pathway bifurcation; Ref. [14] studies a two-variable ODE model consisting of CI and Cro, and adds two independent Gaussian and white noise sources to the equations in order to generate a potential landscape with a complicated mathematical transformation; Ref. [12] uses the stochastic simulation algorithms developed from a master equation, which is generally considered to be the most realistic method of coupled chemical reaction simulation. All these succeed in providing the correct time sequence of the developmental pathways, as well as the probabilities of lysogeny and the genetic switch.

Following recent trends, here we study a different stochastic model to describe the dynamics of the network that regulate the lysis-or-lysogeny choice of coliphage lambda. In this work, we use a Markov chain stochastic model to simulate the behavior of the bacteriophage, according to which all the possible protein states are visited and tracked along their temporal evolutionary paths. Compared to former stochastic models, the disadvantage of the Markov chain model lies in its logiclike dynamics, which makes it impossible to describe the developmental pathway on a real time scale. But this simplicity also leads to advantages. Many global characteristics of the system in phase space can be studied, and it may become possible to define a pseudoenergy for the system. One of the purposes of this paper is to show these advantages. Our results show that all the attractors in the system represent biological ending states that really exist in the life period of coliphage lambda; the biological paths of lytic and lysogenic growths attract many protein states, making themselves attracting trajectories; and the functions of the network tolerate mutational perturbations, showing the robustness of the network. All these features are consistent with the findings of Li *et al.* in their study of the budding yeast cell-cycle network [2]. In addition, the stochastic switch from lysogenic to lytic growth is also revealed in the model. In the analysis, we show a pseudoenergy to the network in order to describe this stochastic process. We found that the energy

\*Electronic address: qi@pku.edu.cn

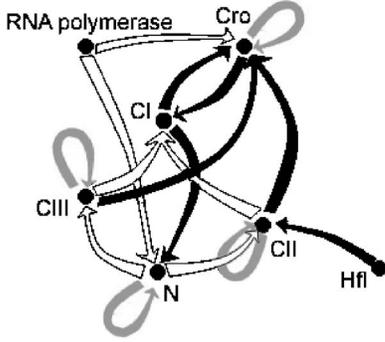


FIG. 1. The network governing the choice between lytic and lysogenic growths of coliphage lambda. Each black node represents a protein; a white arrow presents “activation,” a black arrow means “inhibition.” The arrows with two shafts sharing the same pointed head stand for “collaborative activation/inhibition.” The self-pointed gray ones signify “degradation.” The empirical functions represented by the arrows include gene regulations and protein interactions.

distribution of the network has a transitionlike behavior as randomness increases.

In the next section, we describe the behavior of coliphage lambda and the mechanism that controls its lytic and lysogenic growths. In Sec. III, we build our Markov chain model and describe the method of our numerical simulation. In Sec. IV, we present our simulation results and analysis. In Sec. V, we conclude our report with a brief discussion.

## II. COLIPHAGE LAMBDA

Coliphage lambda is a typical species of temperate bacteriophage. It is described as “temperate” because its life circle involves two different routes, known as lytic and lysogenic growths. The former can be commonly seen in other phages, characterized by lysing of the infected cell after the phage DNA is wildly replicated and assembled with newly expressed protein shells. Usually more than 100 phages come from a single host cell. The latter type of growth is special, because in this route the phage avoids lysing the cell. Instead, it integrates its DNA into the genome DNA of its host. In this case, the phage DNA is called a prophage and the strain has entered a lysogenic state. The prophage can replicate only when the host DNA is replicated. Under certain stimuli, the prophage can be released from the host DNA and start lytic growth again [5,16,17].

The control network of lytic and lysogenic growths of coliphage lambda is shown in Fig. 1 (proteins that do not affect others have been deleted). It can be generalized as follows: First, protein N and Cro are expressed from the phage DNA (RNA polymerase). N can contribute to the expression of protein CII, CIII, O, P, and Q. The Hfl gene of the host produces Hfl proteinase that hydrolyzes CII. In this case, the phage DNA will go through lytic growth. Under certain conditions, such as lack of nutrition or high value of multiplicity of infection (MOI), there is insufficient Hfl proteinase or its activity is restrained. As a result, CII and CIII can work together to start transcription promoter  $P_E$  of the

phage DNA and establish the expression of protein CI, which prohibits the expression of N and Cro. Specifically, the function of CIII is to protect CII, which activates  $P_E$ . CII has almost no activity without CIII, so they can be regarded as cooperators. Highly strengthened transcription of  $P_E$  is an obstruction to the expression of Cro, because the RNA polymerase moves in the opposite direction and a mRNA with the antisequence of Cro mRNA occurs. As a result, CII and CIII cooperate to inhibit Cro expression. Cro can close the production of CI when normally expressed in lytic growth. Hence Cro and CI act as competitors. The function of CII is not only to help express CI, but also to block the yield of Q and to activate the production of integrase, which inserts the phage DNA into the host DNA, making it a prophage. In this state, all transcription is stopped except for those of CI, as it has a special maintenance mechanism. This is the end of lysogenic growth [17–20].

## III. THE MARKOV CHAIN MODEL

In order to understand the global behavior of the control network, we use a discrete model to study the dynamics of the system. In our model, a node  $j$  has only two states:  $S_j = 1$  or  $S_j = 0$ , representing the active or inactive state of the gene (or the protein). The six nodes in the network shown in Fig. 1, namely, Cro, CI, CII, CIII, N, and Hfl, are represented by variables  $(S_1, S_2, \dots, S_6)$ , respectively. Dynamic rules are defined as follows: Define an input function for the  $i$ th incoming arrow of node  $j$  to be  $I_{ij}(t)$  ( $i=1, 2, \dots, n_j$ ) (node  $j$  has  $n_j$  incoming arrows): If it is a white arrow deriving from node  $p$ , then  $I_{ij}(t) = k_w S_p(t)$ ; if it is a black arrow deriving from node  $p$ , then  $I_{ij}(t) = -k_b S_p(t)$ ; for a white arrow deriving from nodes  $p$  and  $q$ ,  $I_{ij}(t) = k_w S_p(t) S_q(t)$ ; for a black arrow deriving from nodes  $p$  and  $q$ ,  $I_{ij}(t) = -k_b S_p(t) S_q(t)$ ; if it is a self-pointed gray arrow,  $I_{ij}(t) = -k_g S_j(t)$ . The  $k$  values here are positive input parameters. The probability that a protein has a certain state in the next time step is determined by the equations below:

$$\text{Prob}[(S_1(t+1), \dots, S_6(t+1)) = \vec{\sigma} | (S_1(t), \dots, S_6(t))] =$$

$$= \prod_{j=1}^6 \text{Prob}[S_j(t+1) | (S_1(t), \dots, S_6(t))], \quad (3.1)$$

where

$$\begin{aligned} \text{Prob}[S_j(t+1) = \delta | (S_1(t), \dots, S_6(t))] &= \\ &= \frac{\exp\left(\beta(2\delta - 1) \sum_i I_{ij}(t)\right)}{\exp\left(\beta \sum_i I_{ij}(t)\right) + \exp\left(-\beta \sum_i I_{ij}(t)\right)}, \end{aligned} \quad (3.2)$$

if  $\sum_i I_{ij}(t) \neq 0$ ,  $\delta \in \{0, 1\}$ ;

$$\text{Prob}[S_j(t+1) = S_j(t) | (S_1(t), \dots, S_6(t))] = \frac{1}{1 + e^{-\alpha}}, \quad (3.3)$$

if  $\sum_i I_{ij}(t) = 0$ . The positive numbers  $\alpha$  and  $\beta$  are parameters characterizing randomness. Notice that when  $\beta=0, \alpha=0$ , the network dynamics is a random process; when  $\beta, \alpha \rightarrow \infty$ , it recovers to the deterministic model of Li *et al.* [2]. Equation (3.2) has a number of forerunners in the literature on neural networks [21].

In the simulation, the state of RNA polymerase is fixed at 1 to ensure an everlasting positive input to N and Cro. Hence the system has a total of 64 ( $2^6$ ) possible states. When neither  $\alpha$  nor  $\beta$  approaches  $+\infty$ , any of the 64 states has a positive probability to become any other state in the next time step. According to theories of the Markov chain, if the above condition is satisfied, there exists a stable probability distribution:

$$\Pi = (\pi_0, \pi_1, \dots, \pi_{63}),$$

$$\pi_s = \lim_{r \rightarrow +\infty} p_{\sigma s}(r), \quad \sigma, s \in \{0, 1, \dots, 63\}. \quad (3.4)$$

$p_{\sigma s}(r)$  here is the  $r$ -step transition probability from the initial state  $\sigma$  to the target state  $s$ . In other words, when  $r$  is big enough, the probability that the system stays at  $s$  is almost independent of the start position  $\sigma$ .

In the simulations, we assume the values of  $k_w, k_b$ , and  $k_g$  are the same for all arrows, and keep them fixed at  $k_w=1, k_b=1$ , and  $k_g=0.1$ . Our simulations show that these control parameters do not qualitatively influence the dynamic behavior of the system. The only control parameters important to the behavior of the system are the “temperature” of the network  $\alpha$  and  $\beta$ , which reflect randomness of the dynamics. In each simulation, we start the system at every possible initial state and let it evolve 100 000 time steps to attain a stable probability distribution of each state. We then let the system evolve another 100 000 time steps to calculate the stable distribution.

#### IV. SIMULATION RESULTS

The simulation resulting from different sets of parameters  $\alpha$  and  $\beta$  are shown in Fig. 2. Defining the pure probability flux from state  $i$  to  $j$  to be  $\pi_i p_{ij} - \pi_j p_{ji}$ , we use the biggest outgoing pure flux to represent the trajectories or state flows of the system. Actually, since the system is ergodic, there are fluxes between any two states, but only the biggest outgoing pure flux of each state is displayed in Fig. 2 to show the main trend of the temporal evolution. In the figure, each node stands for a system state, and the directed edges represent state flows. In Fig. 2(a), the size of the nodes and the thickness of the edges are proportional to the logarithm of the total traffic flow passing through them; in Fig. 2(b), they are proportional to the square root of the total flow. The dark trajectories are the biological trajectories of lytic (2) and lysogenic (3) growths. The three dark nodes represent biological stationary states (1: lysogen with  $Hfl=1$ ; 2: lysis; 3: lysogeny with  $Hfl=0$ ). The out-going arrow from dark node 2 is not drawn because that point is the end of the lytic choice. The phage lambda DNA will soon express a series of proteins that are not included in the regulatory network used here.

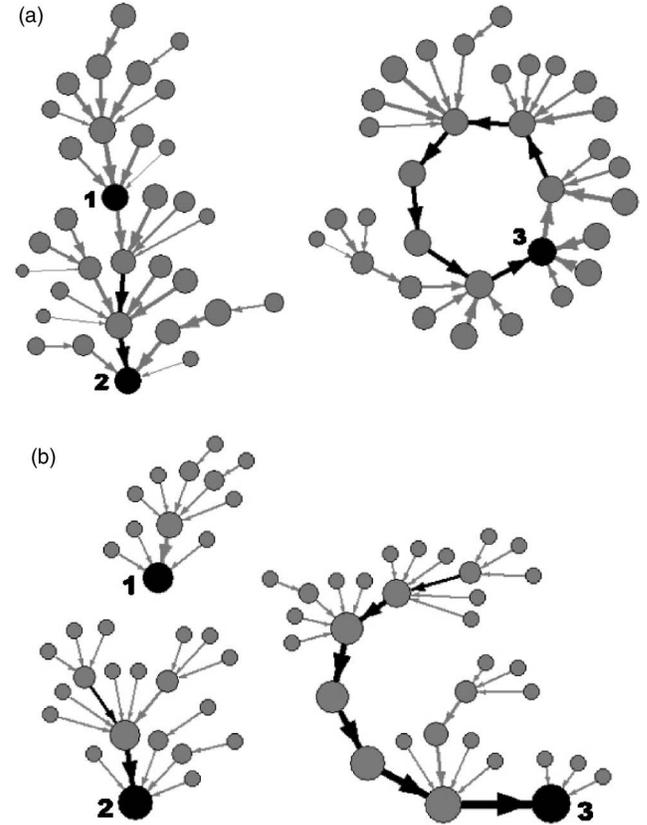


FIG. 2. The protein state temporal evolution trajectories, which is represented by the biggest outgoing pure flux. Parameters: (a)  $\alpha=4, \beta=3$ ; (b)  $\alpha=\infty, \beta=\infty$ .

As shown in Fig. 2, when the control parameters  $\alpha$  and  $\beta$  are beyond certain values, the biological state pathways and biological stationary states experience great attractions, i.e., all the states converge to a few fixed points corresponding to biological steady states; and the biological trajectories, which are the real pathways of the genetic growths, absorb many state nodes before they flow to the fixed points. Compare Fig. 2(a), where  $\alpha=4, \beta=3$ , with Fig. 2(b), where  $\alpha=\beta=\infty$ , one notices that the stochastic model reflects not only the attractions of the biological states and pathways, but also both the possibilities that the lysogenic ending state with  $Hfl=0$  (dark node 3) can flow to the start point of lysogenic growth, and that lysogen with  $Hfl=1$  (dark node 1) can flow to the start point of lytic growth. These are consistent with the fact that lysogen can stochastically switch to the very beginning of the growth choice.

To test the robustness of this lambda regulatory network, the effects of two types of perturbations on the network are considered: mutations and noisy environments. For mutational perturbation, three types of perturbations are introduced in the deterministic model ( $\alpha \rightarrow \infty, \beta \rightarrow \infty$ ): arrow additions, arrow deletions, and arrow-color switches (white to black or black to white). For arrow additions, the two host

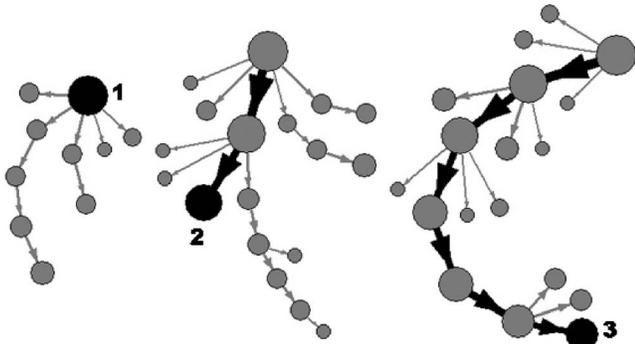


FIG. 3. Perturbation analysis. The sizes of the nodes and the thickness of the edges are proportional to the square root of the traffic flows passing through them.

proteins (RNA polymerase and Hfl) are allowed to be added only with out-pointed arrows. Shared-arrow additions are not considered, and only one perturbation is conducted each time.

The three pictures in Fig. 3 show the average flow of the temporal state evolutions in perturbed networks starting from the biological beginning states of descended lysogeny (1), lytic growth (2), and lysogenic growth (3), respectively. To make the figures clearer, only the greatest incoming flux of each node is displayed. One observes that, on average, the perturbed networks are less stable than the original one: The attractors (dark nodes) become leaky; the probability of leaving these states increases. Most of the states in the system, however, still converge to these attractors. For the descended lysogeny process (1), we conducted 14 arrow deletions, 36 arrow additions, and 10 color switches. The percentages of the perturbed networks that still evolve to the biological end are 86, 75, and 70 %, respectively. For the lytic growth process (2), we conducted 14 arrow deletions, 35 arrow additions, and 10 color switches. The function-preserved percentages are 71, 69, and 50 %, respectively. For the lysogenic

growth process (3), we conducted 13 arrow deletions, 35 arrow additions, and 9 color switches. The function-preserved percentages are 23, 60, and 33 % respectively. The difference in the numbers of possible perturbations derives from the fact that a few perturbations may lead to the occurrence of circulated state pathways or limit cycles, which are not included in the analysis. These results show that the network functions are largely preserved, because in most cases, the states still evolve to the biological fixed points (dark nodes) through the biological pathways (dark trajectories).

To investigate the effect of the environmental fluctuations, we study the dynamical behavior of the network as a function of parameters  $\alpha$  and  $\beta$ , which reflect the randomness of the network dynamics. As  $\alpha$  and  $\beta$  are parameters characterizing randomness of different parts of the system, they should increase or decrease simultaneously with environmental changes. In this study, the relation between  $\alpha$  and  $\beta$  is set to be  $\beta = K\alpha$ , where  $K$  is a positive constant. We define a pseudoenergy of each system state as follows:

$$E_s = \frac{-\ln(\pi_s)}{\beta}, \quad s \in \{0, 1, \dots, 63\}. \quad (4.1)$$

Here, the concept of pseudoenergy is borrowed from equilibrium statistical physics. We hypothesize that the biological stationary states, i.e., the big attractors, will be at the minima in the energy landscape when the level of randomness of the system is not very high, so that the three stationary states act as potential traps. The reason the system could escape from the traps and switch to other kinds of growths is the stochastic force or the intrinsic noise.

Our simulation shows that the energy distribution of all the states change as a function of  $\alpha$ , as shown in Fig. 4 (white poles). One observes that with the increase of  $\alpha$ , the energy distribution goes through a fast change, similar to a phase transition in statistical physics. Before the critical point ( $\alpha_c = 6$ ), the energy distribution is Gaussian-like, which

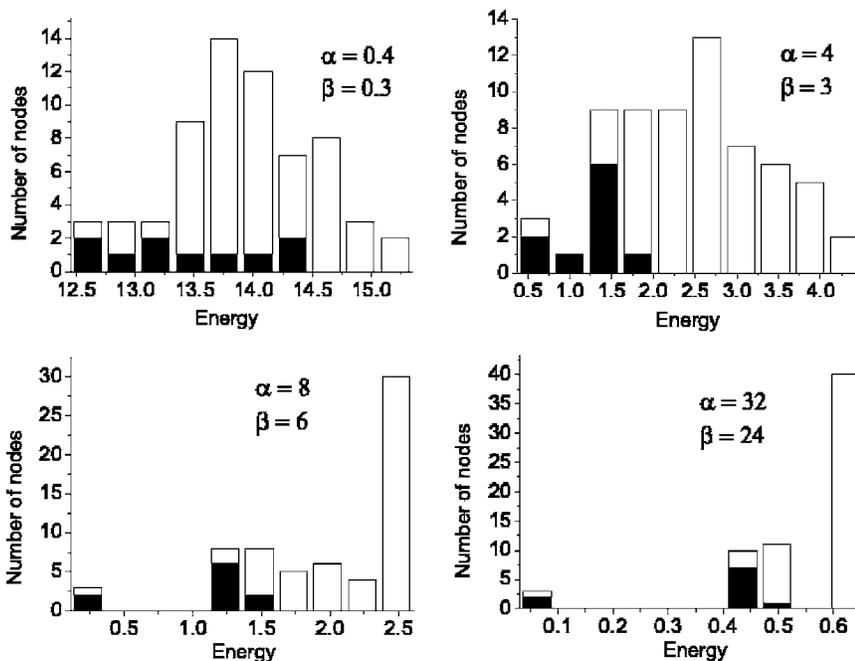


FIG. 4. Energy distributions with different levels of randomness. Parameters:  $k_w = k_b = 1, k_g = 0.1, K = 0.75, \alpha = 0.4, 4, 8, 32$ . The white poles (or bars) signify all the 64 states, while the black poles (or bars) stand for the 10 states on the biological attracting trajectories.

indicates that the energy is more or less evenly distributed in the phase space; the system visits each state with equal probability. After the critical point, the number of nodes with high energy suddenly increases, which means that most states tend to be unstable, and the system will rapidly converge to a few states with low energy, including the three stationary states and states on the attracting trajectories. The black poles (or bars) in Fig. 4 display the energy distribution of the states on the biological attracting trajectories. It is shown that when the level of randomness is high, this distribution is even; as the level of randomness decreases, the states on the biological trajectories gain lower potentials than most of the other states, symbolizing the formation of the attracting trajectories. It can be derived from Eqs. (3.2) and (3.3) that if  $\alpha=0$ , all the protein states will share the same energy level as the network becomes a sheer random one.

An interesting question is whether such a transitionlike behavior is unique to biological networks; could it also be observed in structure-random networks, in which the activation and prohibition relations between proteins are randomly generated? Our simulations with random networks show that the structure-random networks also give this transitionlike behavior.

## V. DISCUSSION

This work demonstrates the stability and robustness of the lysis-or-lysogeny regulatory network, which plays a critical role in the life cycle of coliphage lambda. The results show that the biological stationary states are big attractors and the paths of lysogenic and lytic growths are attracting trajectories. The network functions are mostly preserved when the network is perturbed. Similar dynamic properties were seen in the cell-cycle network of budding yeast, fission yeast, and frog egg [2].

The Markov chain method based on a Boolean model has certain advantages. It ensures that many global characteris-

tics of the system in phase space can be studied, unlike general methods in nonlinear analysis which are valid only around the fixed points in phase space, providing only local information. It is possible that a pseudoenergy can be defined for more complex system, works on this line of thinking is underway in our group.

The pseudoenergy distribution of the system states seems to have a transitionlike behavior as randomness in the model is increased. It is reasonable to suppose that these properties may be common for most regulatory networks governing important genetic processes of organisms. It should be noticed that the results of our simulation and the simulation of cell-cycle network dynamics [22] show that the pseudoenergy behaves similar to “free-energy” in equilibrium statistical physics, meaning in the steady state, the system is at the minimum pseudoenergy; if there are more than one minima, a control parameter change can switch the system from one state to the other. However, whether this concept can be applied to more general network dynamics is still an open question. For example, if the system has a limit cycle, this method will be invalid. How to construct a more general pseudoenergy to translate a general network dynamic system to a potential system is a subject of study in our group.

Future research will surely provide us with more insight into how the topology of regulatory networks can affect genetic functions. With the help of a more detailed and extended regulatory network model, there may be better methods to analyze the probability and stability of mutations and to predict the trend of evolutions.

## ACKNOWLEDGMENTS

We thank F.T. Li and M.P. Qian for helpful discussions. This work is supported by the Department of Science and Technology of China (2003CB715900) and the Chinese Natural Science Foundation. S.Y.W. and Y.P.Z. contribute equally to this work.

- 
- [1] R. Albert and A. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).  
 [2] F. T. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang, *Proc. Natl. Acad. Sci. U.S.A.* **101**, 4781 (2004).  
 [3] R. Thomas, A. M. Gathoye, and L. Lambert, *Eur. J. Biochem.* **71**, 211 (1976).  
 [4] G. K. Ackers, A. D. Johnson, and M. A. Shea, *Proc. Natl. Acad. Sci. U.S.A.* **79**, 1129 (1982).  
 [5] M. A. Shea and G. K. Ackers, *J. Mol. Biol.* **181**, 211 (1985).  
 [6] J. Reinitz and J. R. Vaisnys, *J. Theor. Biol.* **145**, 295 (1990).  
 [7] H. H. McAdams and L. Shapiro, *Science* **269**, 650 (1995).  
 [8] J. D. Chung and G. Stephanopoulos, *Chem. Eng. Sci.* **51**, 1509 (1996).  
 [9] H. McAdams and A. Arkin, *Proc. Natl. Acad. Sci. U.S.A.* **94**, 814 (1997).  
 [10] H. H. McAdams and A. Arkin, *Annu. Rev. Biophys. Biomol. Struct.* **27**, 199 (1998).  
 [11] J. Vohradsky, *J. Biol. Chem.* **276**, 36168 (2001).  
 [12] A. Arkin, J. Ross, and H. H. McAdams, *Genetics* **149**, 1633 (1998).  
 [13] E. Aurell and K. Sneppen, *Phys. Rev. Lett.* **88**, 048101 (2002).  
 [14] X. M. Zhu, L. Yin, L. Hood, and P. Ao, *Funct. Integr. Genomics* **4**, 188 (2004).  
 [15] T. H. Tian and K. Burrage, *J. Theor. Biol.* **227**, 229 (2004).  
 [16] A. D. Johnson, A. R. Poteete, G. Lauer, R. T. Sauer, G. K. Ackers, and M. Ptashne, *Nature (London)* **294**, 217 (1981).  
 [17] M. Ptashne, *A Genetic Switch: Phage  $\lambda$  Revisted* (Cold Spring Harbor Laboratory Press, New York, 2004).  
 [18] I. Herskowitz and D. Hagen, *Annu. Rev. Genet.* **14**, 399 (1980).  
 [19] H. Echols, *Science* **233**, 1050 (1986).  
 [20] D. I. Friedman, *Curr. Opin. Genet. Dev.* **2**, 727 (1992).  
 [21] J. Hertz, A. Krogh, and R. G. Palmer, *Introduction to the Theory of Neural Computation* (Addison-Wesley, Reading, MA, 1991).  
 [22] Y. P. Zhang, M. P. Qian, Q. Ouyang, F. T. Li, and C. Tang, report, 2005 (unpublished).