ACS SyntheticBiology

Research Article

pubs.acs.org/synthbio

Landscape Topography Determines Global Stability and Robustness of a Metabolic Network

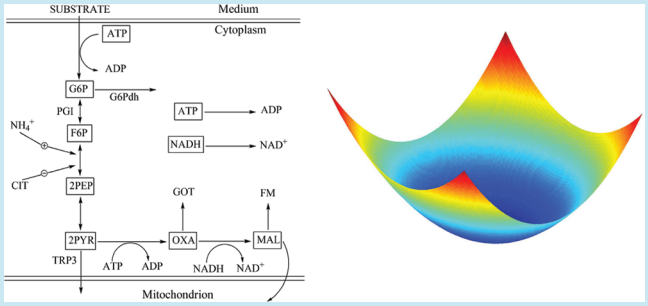
Chunhe Li,[†] Erkang Wang,*[,†] and Jin Wang*[,†,‡]

[†]State Key Laboratory of Electroanalytical Chemistry, Changchun Institute of Applied Chemistry, Chinese Academy of Sciences, Changchun, Jilin 130022, P. R. China

[‡]Department of Chemistry and Physics, Stony Brook University, Stony Brook, New York 11794, United States

Supporting Information

ABSTRACT: Metabolic networks have gained broad attention in recent years as a result of their important roles in biological systems. However, how to quantify the global stability of the metabolic networks is still challenging. We develop a probabilistic landscape approach to investigate the global natures of the metabolic system under external fluctuations. As an example, we choose a model of the carbohydrate metabolism and the anaplerotic synthesis of oxalacetate in *Aspergillus niger* under conditions of citric acid accumulation to explore landscape topography. The landscape has a funnel shape, which guarantees the robustness of system under fluctuations and perturbations. Robustness ratio (*RR*), defined as the ratio of gap between lowest potential and average potential versus roughness measured by the dispersion or square root of variations of potentials, can be used to quantitatively evaluate the global stability of metabolic networks, and the larger the *RR* value, the more stable the system. Results of the entropy production rate imply that nature might evolve such that the network is robust against perturbations from environment or network wirings and performs specific biological functions with less dissipation cost. We also carried out a sensitivity analysis of parameters and uncovered some key network structure factors such as kinetic rates or wirings connecting the protein species nodes, which influence the global natures of the system. We found there is a strong correlation between the landscape topography and the input-output response. The more stable and robust the metabolic network is, the sharper the response is.



KEYWORDS: *probabilistic landscape, global stability, metabolic network, fluctuations, robustness ratio*

Developing quantitative and global methodologies to investigate large-scale biochemical networks has become one of the important questions in the post-genomic era. Metabolic networks, as dynamically regulated, complex interactive nonlinear systems, have been studied extensively in recent years.[1−6] However, how to quantify the global stability of metabolic networks is still challenging. So far there are limited studies from the physical viewpoints to investigate why metabolic networks are so robust and how they perform their biological functions.

In bulk dynamics, deterministic chemical reaction equations can be used to describe the local dynamics of metabolic networks. However, the global natures of the system are hard to address using this method. For a high-dimensional network with huge state space, a challenging job is to understand how seemingly infinite number of genotypes can produce a finite number of functional phenotypes. A probabilistic landscape approach may provide a possible route for this, because according to landscape theory different states correspond to different probabilities to appear. The functional state should possess higher probability of appearance or lower potential energy, while unfunctional states should have lower probability or higher potential.[7−12] In the cell, there are two sources of noise including intrinsic noise from statistical fluctuations of the finite molecular number and external noise from highly dynamical and inhomogeneous environments.[13−19] Therefore, one should study chemical reaction networks in fluctuating conditions to model more realistic cellular environments. The extrinsic noise can be introduced by external random force,[20] and the corresponding probabilistic evolution follows a diffusion equation. For intrinsic fluctuations, one can explore master equation formalism,[21] which gives the evolution of probability due to random natures of the system in terms of numbers.

However, a master equation or diffusion equation is difficult to solve in high-dimensional state space. The degrees of freedom grow exponentially with the number of the components (different species of proteins in the metabolic networks). We will employ a self-consistent mean field approximation to split the components to effectively reduce the dimensions of the system, and make it tractable for computations. In this way, the degree of freedom reduces from N^M to $N{\cdot}M$, where N is the concentration intervals (how many

Received: March 23, 2012
Published: May 4, 2012

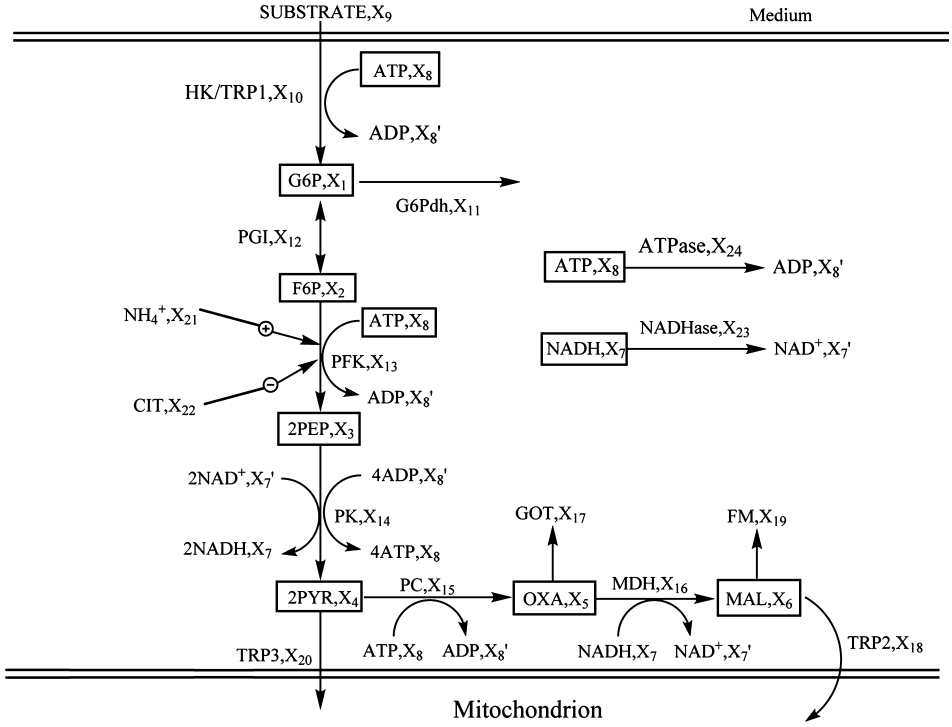


Figure 1. Metabolic network model of carbohydrate metabolism in *Aspergillus niger* in the condition of citric acid production. Dependent variables are numbered 1−8: G6P, X_1, glucose-6-phosphate; F6P, X_2, fructose-6-phosphate; PEP, X_3, phosphoenol pyruvate; PYR, X_4, pyruvate; OXAL, X_5, oxaloacetate; MAL, X_6, malate; NADH, X_7, NADH; NAD, X'_7, NAD; ATP, X_8, ATP; and ADP, X'_8, ADP. Independent variables are numbered 9−24: Substrate, X_9, substrate; HKTRP1, X_{10}, hexokinase-substrate transport step; G6P-DH, X_{11}, glucose-6-phosphate dehydrogenase; PGI, X_{12}, phosphoglucose isomerase; PFK, X_{13}, phosphofructokinase; PK, X_{14}, pyruvate kinase; PC, X_{15}, pyruvate carboxylase; MDHAL, X_{16}, malate dehydrogenase; GOT, X_{17}, aspartate aminotransferase; TRP2, X_{18}, malate transport; FM, X_{19}, fumarase; TRP3, X_{20}, pyruvate transport; NH_4^+, X_{21}, ammonium; Cit, X_{22}, citrate; DHase, X_{23}, dehydrogenases; and ATPase, X_{24}, ATPases.

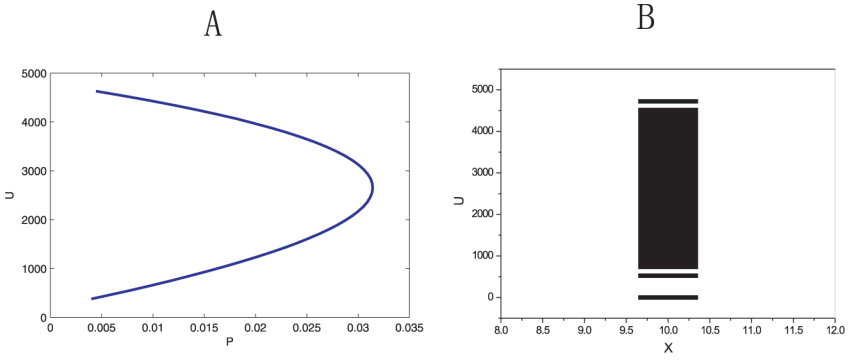


Figure 2. (A) Distribution and (B) spectrum of the potential energy U.

segments of each concentration variable) and M is the number of variables (how many different kinds of proteins).

The solutions of diffusion equation or master equation provide information on probabilistic evolution. We will explore the global natures of the network from physical perspectives, in terms of the potential landscape[11,12,22,23] closely linked to the steady state probability. To illustrate our ideas we studied a metabolic network as an example,[24,25] which is a model of the carbohydrate metabolism and the anaplerotic synthesis of oxalacetate in *Aspergillus niger*, under conditions of citric acid accumulation. According to the underlying potential landscape of the network, we will investigate the global stability quantitatively under the fluctuations by quantifying landscape topography and calculating the robustness ratio (RR). We will also do the sensitivity analysis of parameters to uncover some key network structure factors such as kinetic rates or wirings connecting the protein nodes, which determine the global stability and the function of the metabolic system.

RESULTS AND DISCUSSION

Figure 1 gives an illustrative wiring diagram of the metabolic network with 8 variables of protein concentration. In terms of the self-consistent mean field approximation, we can obtain the steady state probability distributions of the 8 variables of protein concentrations giving diffusion coefficient D characterizing environmental fluctuation strengths. Then we can infer the potential landscape according to the links between the generalized potential U and the steady state probability $P_{ss}(x)$: $U(\mathbf{x}) = -\ln P_{ss}(\mathbf{x})$.[7−12,22,23,26] For the 8-dimensional network, it is difficult to visualize and analyze the landscape directly.

dx.doi.org/10.1021/sb300020f | *ACS Synth. Biol.* 2012, 1, 229−239

Therefore, we first investigate a lower dimensional projection of it.

Here, for the metabolic network, we first used a constant diffusion coefficient to calculate the results, i.e., the diffusion coefficient matrix $D[x(t)]$ is a constant diagonal matrix. This choice more likely describes the external environmental fluctuations, because generally external fluctuations do not necessarily depend on the internal concentrations of proteins (variable x). In the Supporting Information we also give the results when the diffusion coefficient is concentration-dependent, which quantifies the influence of intrinsic noise. For these two forms of diffusion matrix, we obtained similar results, which are stated in the main text and Supporting Information.

Zero Dimensional Projection of Landscape and Robustness of System. First, we studied the spectrum and distribution of the potential U, since the potential is a multidimensional function in protein concentration x space. Figure 2A is a zero dimensional projection of potential U. We can see that the distribution is approximately Gaussian. The lowest potential energy U is the global minimum of the potential landscape, which is at about the same place (in concentration space) as the deterministic solutions of the averaged chemical reaction equations for the system. Figure 2B illustrates the spectrum of the potential. We can see clearly that the global minimum of the potential is significantly far from the rest of the spectrum or the distribution. There exists a significant gap between the ground state global minimum and the rest of the potential spectrum. Since the ground state probability is exponentially associated with the potential, a large gap in potential implies large discrimination in populations of ground state against the other states. In other words, the basin of attraction has much larger weights than any other states of the system. This can guarantee the stability of the system.

From the underlying potential landscape spectrum, we can further explore the robustness of the system against environmental perturbations and fluctuations. Here, we define robustness ratio RR for the network as the ratio between the gap δU and roughness ΔU of the underlying potential landscape. δU is the difference between the global minimum U_{min}, and the average of U, $\langle U\rangle$, and ΔU is the spread or the half width of the distribution of U. So $RR = \delta U/\Delta U$. δU is a measure of the bias or the slope toward the global minimum of the potential landscape, whereas ΔU is a measure of the averaged roughness or the local trapping of the potential landscape. The larger the robustness ratio is, the bigger the gap compared with roughness and the probability of the basin of attraction are, and so the more stable the system is. So we can see the robustness ratio $RR = \delta U/\Delta U$ quantifies the topography of the underlying landscape.

We obtained the landscape results separately for different diffusion coefficients D, characterizing the external fluctuations. Thus, the effects of the external fluctuations on the system and associated landscape can be explored. Figure 3A shows the robustness curve when external noise D is changed. We can see clearly that RR decreases when the diffusion coefficient D increases, which means that larger external fluctuations destroy the stability of the system.

For the perturbations of chemical reaction rate constants, they effectively change the strengths of the wirings in the network. We changed the parameters of the equations in terms of a probability distribution with a mean of the unperturbed rate, k_u, and a standard deviation $\sigma = lp \cdot k_u$; here lp represents the perturbation level applied to the system. For every

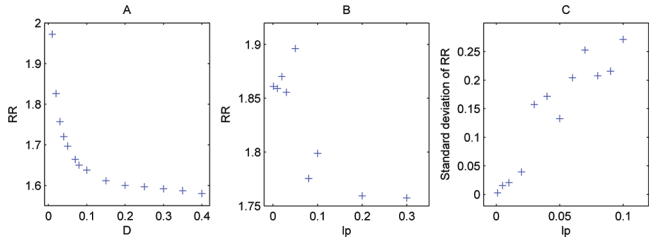


Figure 3. Robustness rate ($RR = \delta U/\Delta U$; δU is the difference between the global minimum U_{min} and the average of U, $\langle U\rangle$, and ΔU is the spread or the half width of the distribution of U) versus (A) external noise D and (B) perturbation level of wiring strengths lp. (C) Standard deviation of RR versus perturbation level of wiring strengths lp.

perturbation level lp, the corresponding landscape can be acquired. Therefore, by changing lp we can investigate the effects of perturbations of chemical reaction rate constants on the landscape and therefore the robustness of the system.

From Figure 3B we can see that robustness decreases when the perturbation level lp increases. This implies that when certain perturbations are given to parameters the metabolic system is inclined to be less stable, and therefore the system with current parameters might be the more stable one. The parameters of the metabolic network we have chosen are from Torres' work,[24] which gave the parameter values of the model and the corresponding references where these parameters were obtained directly by experiments or by estimations according to the Michaelis−Menten rate equations.[27] Therefore, the parameters we have chosen for the model have experimental justifications, and our results in terms of the landscape of the network show that the current parameter values might be those making the system most stable, which is reasonable from the view of evolutional selection.

In addition, we also explored the standard deviation of RR when perturbation level lp was changed. Figure 3C shows that the standard deviation of RR increases as lp increases. This implies that larger variances or changes of the connection strengths of wirings of the network lead to less stability. Therefore, large RR guarantees the robustness of the system against perturbations in the underlying chemical reaction rates characterizing the wiring strengths of the network as well as the external fluctuations.

One Dimensional Projection of Landscape: *U* versus *Q* and *U* versus rmsd. Figure 4 shows the one-dimensional projection of U on the overlapping order parameter Q with respect to the global minimum. Here Q is defined as $Q = (\sum_i^N x_i x_i^{min})/(|x||x^{min}|)$, so that we can keep track of the degree of "closeness" or overlap between an arbitrary state x to the global minimum state x_{min} in the state space of the protein concentrations. $Q = 1$ represents the global minimum state and $Q = 0$ represents the decorrelated states with no overlap with the global minimum. Figure 4A shows U versus Q at different diffusion coefficient D. From Figure 4A, we can see a downhill slope of the potential U in Q toward the global minimum U_{min}. The potential U reaches to the global minimum U_{min} when $Q = 1$. This shows a funnel of U along Q toward the global minimum of the potential landscape. In addition, we can see that the slop of the curve decreases as D characterizing external noise increases. This shows that with external noise increased, the slope of the funnel decreases and the system is less stable.

We also calculated the results of U versus root mean squared distance ($rmsd = [\sum_i^N i(x_i - x_i^{min})^2]^{1/2}$). Here rmsd represents

dx.doi.org/10.1021/sb300020f | ACS Synth. Biol. 2012, 1, 229−239

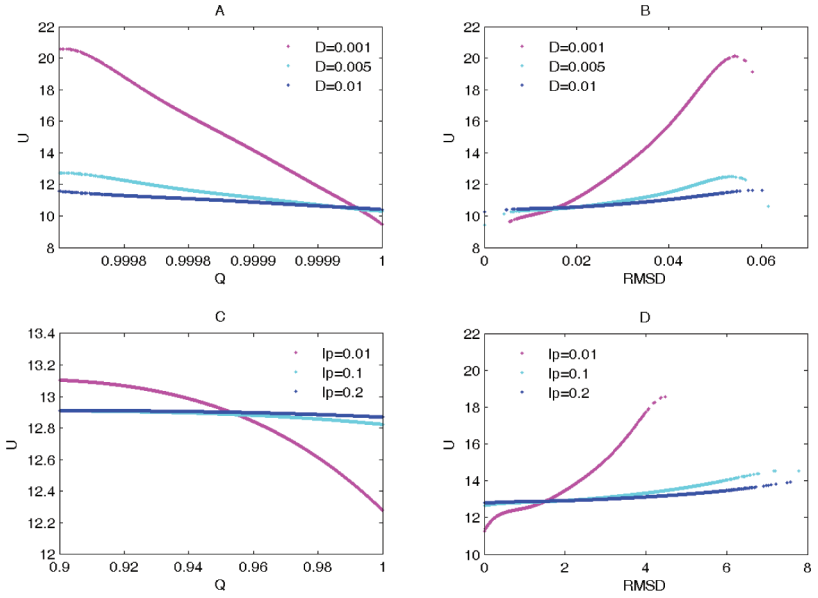


Figure 4. (A) Q ($Q = (\sum_i^N x_i x_i^{\min})/(|x||x^{\min}|)$, representing the degree of "closeness" or overlap between an arbitrary state x to the global minimum state) versus U and (B) rmsd (rmsd = $(\sum_i^N (x_i - x_i^{\min})^2)^{1/2}$, root mean squared distance, representing the distance between a state point and global minimum in state space) versus U with diffusion coefficient D changed. (C) Q versus U and (D) rmsd versus U with perturbation level of wiring strengths lp changed.

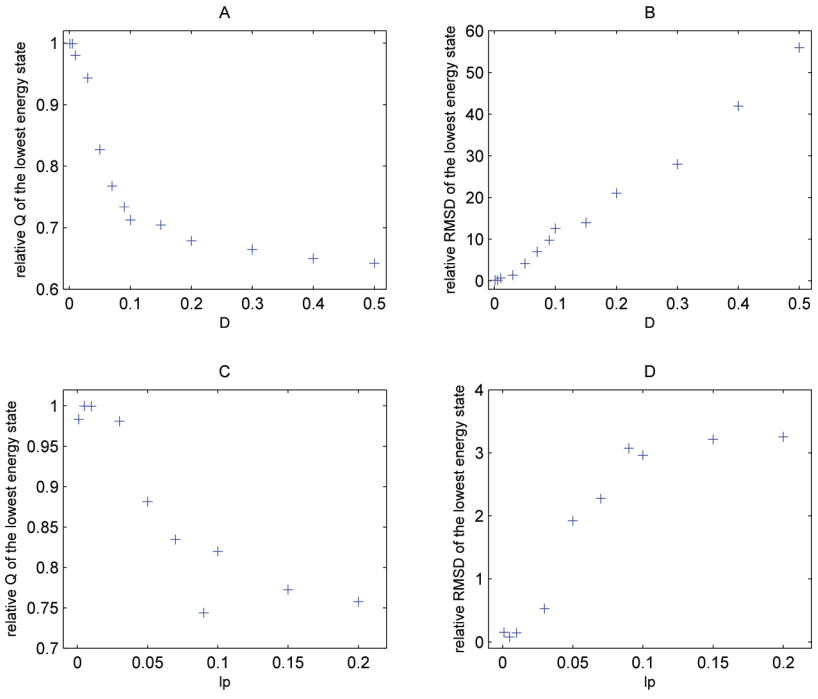


Figure 5. (A) Relative Q and (B) rmsd of the lowest energy state at different diffusion coefficient D. (C) Relative Q and (D) rmsd of the lowest energy state at different perturbation level of wiring strengths lp.

the distance between a state point and global minimum in state space. A large rmsd corresponds to small Q. Figure 4B shows U versus rmsd at different diffusion coefficient D. We can see again a funnel landscape down to the basin of attraction, and the landscape shape is shallower when D increases.

We further varied the perturbation level lp of chemical reaction rates or wiring strengths to explore the change of the slope of the landscape funnel. From Figure 4C,D, we see that increasing perturbation levels lp of rates of wiring strengths of metabolic networks will lead to the landscape slope being less biased or tilted toward the global minimum or the basin of attraction and more flat. So the slope of the landscape can also be a measure for the robustness of the network. Therefore, fewer fluctuations and perturbations lead to a higher slope and more funneled landscape, and the system will be more stable and robust.

Additionally, we explored the relative Q and rmsd of the lowest energy state or the ground state at different perturbations. Here we took original rates or wiring strengths and smallest D as the reference state with which to compare

dx.doi.org/10.1021/sb300020f | ACS Synth. Biol. 2012, 1, 229−239

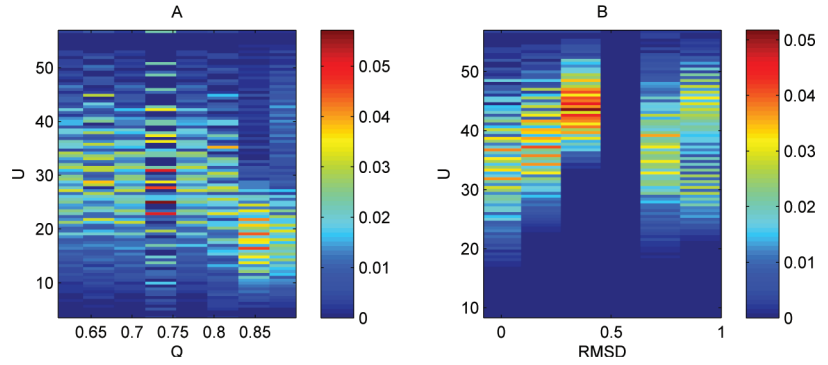


Figure 6. (A) Distribution of U at different Q. (B) Distribution of U at different rmsd.

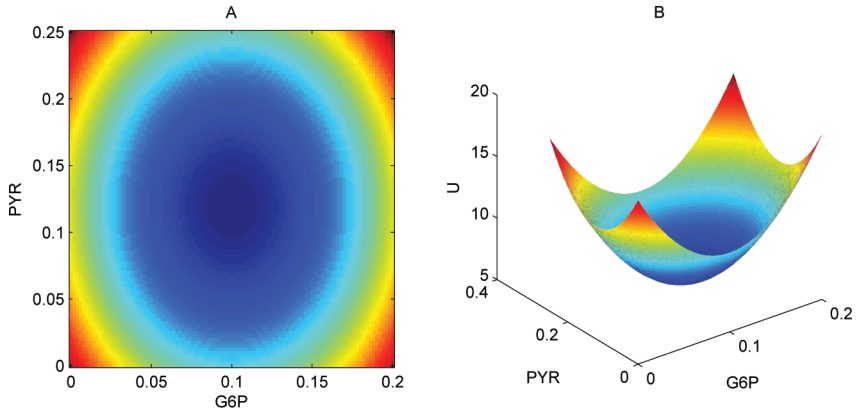


Figure 7. (A) Two-dimensional and (B) 3-dimensional landscape of the metabolic network for the two variables X_1 (G6P) and X_4 (PYR).

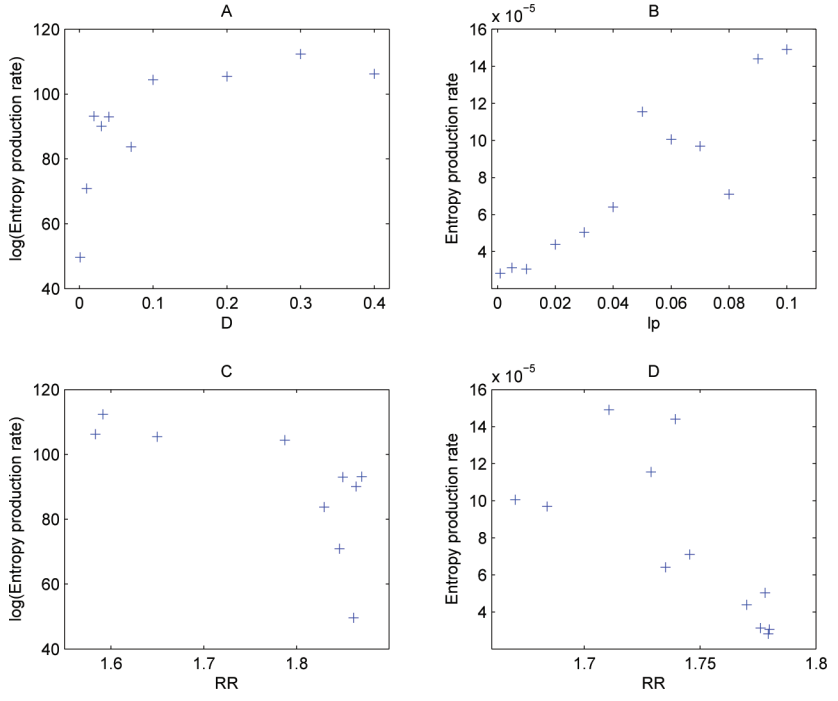


Figure 8. Entropy production rate versus (A) external noise D and (B) perturbation level lp. Entropy production rate versus RR with (C) external noise D changed and (D) perturbation level of wiring strengths lp changed.

others. Figure 5A,B shows the effects of external noise on the relative Q and rmsd. We can see that the relative Q of the ground state decreases and the relative rmsd of the ground state increases when D increases. This shows that at large external noise the ground state becomes further away from the ground state where there is no noise, and the system becomes less stable. Figure 5C,D shows the effects of perturbation level lp of rate parameters or wirings on the relative Q and rmsd. With the

dx.doi.org/10.1021/sb300020f | *ACS Synth. Biol.* 2012, 1, 229−239

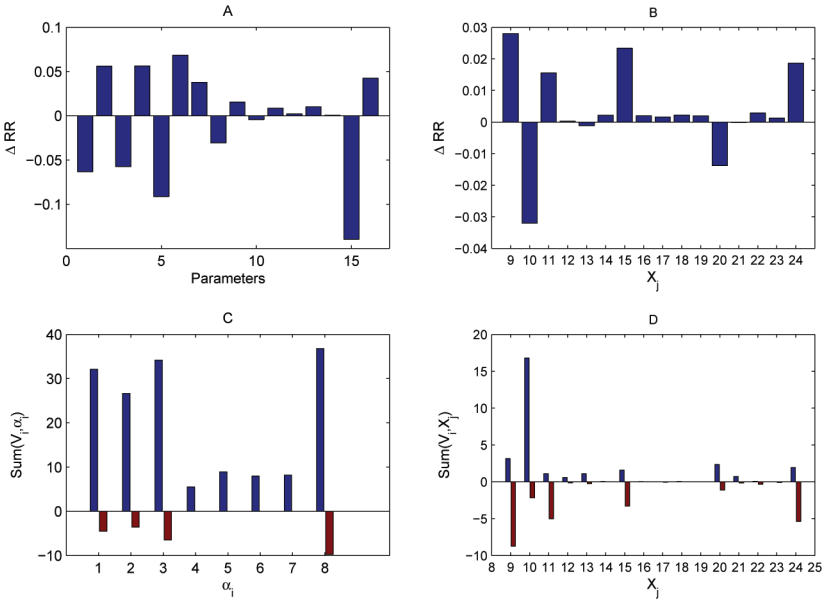


Figure 9. Effects of (A) rate constants and (B) independent variables on the robustness. In panel A, the x axis represents: α_1,β_1, α_2,β_2, ..., α_8,β_8, In panel B, the x axis represents X_9, X_{10}, ..., X_{24}. Magnitudes for every (C) rate constant α_i and (D) independent variable X_j summed over all fluxes V_i. Blue bars represent positive flux, and red bars represent negative flux.

same trend of slope compared with Figure 5A,B, it shows that the system becomes less stable as perturbation level *lp* increases.

Figure 6A shows the distribution of U at different Q, and Figure 6B shows the distribution of U at different rmsd. Blue color represents lower probability, and red color represents higher probability. From the figure, we can see that at every Q and rmsd, the distribution of U is a Gaussian distribution, and when Q is big or rmsd is small, the distribution of U moves toward a lower potential energy state. This is reasonable because a larger Q or smaller rmsd means that the state approaches the ground state and therefore possesses lower potential.

Projected Landscape in Two- and Three-Dimensional Space and Entropy Production Rate. We studied the 2-dimensional and 3-dimensional landscape of the metabolic network. Figure 7 shows the 2-dimensional and 3-dimensional landscape separately projected on variable X_1 (G6P) and X_4 (PYR) for the diffusion coefficient D = 0.001. From the landscape we can see that there is one stable basin of attraction for the metabolic network, and the funneled landscape shape guarantees the stability and robustness of the system.

In addition, we calculated the entropy production rate or dissipation cost for different fluctuations.[28] Figure 8A,B shows that the entropy production rate increases as external noise D or perturbation level *lp* increases. This shows that at larger fluctuations from external environment or internal wirings, the system costs more dissipations. We also investigated the entropy production rate versus *RR* at different fluctuations. Figure 8C,D shows the curve of the entropy production rate versus *RR* separately for changing D and changing *lp*. We can see for both conditions the entropy production rate decreases as the robustness ratio *RR* of the system increases. This shows that the less the variation of rate parameters or wiring strengths or external fluctuations, the more robust the network is and the less the entropy production or dissipations for the network are. This can be important for the network design. This implies that nature might evolve such that the network is robust against internal and environmental perturbations and performs specific biological functions with less dissipation cost. In our study, this is also the equivalent of optimizing the robustness or stability of the network.

Sensitivity Analysis of Parameters and Control of Flux. To discover the key connections or wirings of the metabolic network responsible for the stability of the system, we did sensitivity analysis of parameters. For the chemical reaction rate equations of the metabolic network, each equation has one pair of rate constants corresponding separately to synthesis and degradation reactions, and there are totally 16 rate constants for 8 reactions. By analyzing the influence of these rate constants on the robustness of system, we can obtain which reactions are more important and further which protein elements and parameters are crucial in maintaining the robustness of the system.

Here we gave rate constants a variance $(\Delta k/k)$ to change. We can compute robustness at different $\Delta k/k$ for different rate constants. Figure 9A shows the results of robustness for 16 rate constants in different $\Delta k/k$. We selected the four top important parameters corresponding to the four most crucial reactions. They are the synthesis reactions separately for G6P (X_1), F6P (X_2), PEP (X_3), and ATP(X_8):

$$\begin{aligned} X_8 + X_9 + X_{10} &\rightarrow X_1 \\ X_1 + X_{12} &\rightarrow X_2 \\ X_2 + X_8 + X_{13} + X_{21} + X_{22} &\rightarrow X_3 \\ X_3 + X_7 + X_8 + X_{14} &\rightarrow X_8 \end{aligned} \quad (1)$$

Compared with the diagram in the Figure 1, we can find that for these four reactions the first three of reactions (the synthesis of G6P (X_1), F6P (X_2), PEP (X_3)) are the upstream parts of the whole metabolic pathway, which represent the most important process in process (i), the formation of pyruvate (X_4). These three reactions determine the formation of pyruvate (X_4). For the last reaction, the synthesis reaction of ATP (X_8), it should be very important, because that ATPs appear in many places

dx.doi.org/10.1021/sb300020f | *ACS Synth. Biol.* 2012, 1, 229–239

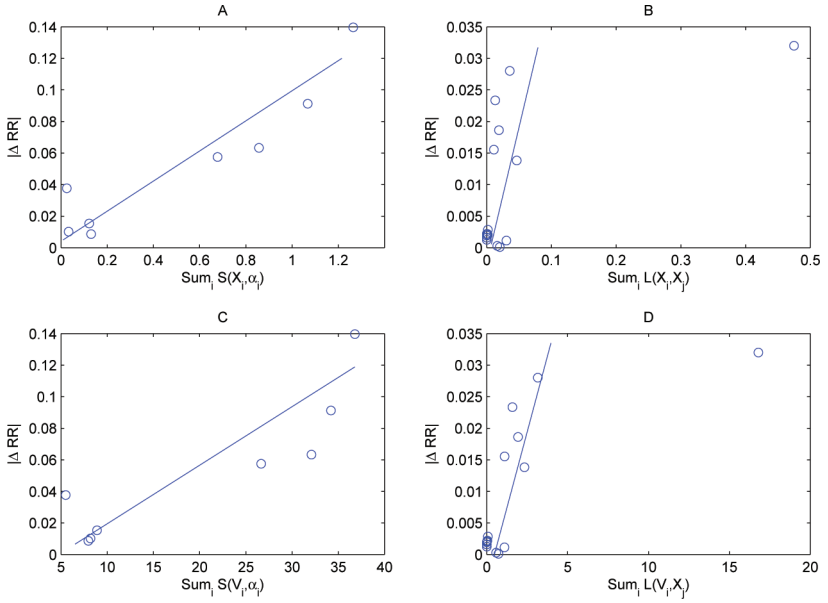


Figure 10. Correlation of Δ*RR* and concentration for (A) sensitivity of parameters and (B) logarithmic gains of independent variables. Correlation of Δ*RR* and flux for (A) sensitivity of parameters and (B) logarithmic gains of independent variables.

and play important roles on the whole metabolic pathway, which we can see clearly in Figure 1.

Figure 9(B shows the effects of independent variables X_j (j = 9, 10, ..., 24) on the robustness and stability of the system. We can find some important independent variables including X_9, X_{10}, X_{11}, X_{15}, X_{20}, X_{24}.

We also computed the influence of rate constants and independent variables separately on the fluxes. Sensitivity with respect to the change of rate constants is defined:[24,25,29]

$$S(X_i, \alpha_i) = \frac{\partial \ln(X_i)}{\partial \ln(\alpha_i)} \tag{2}$$

The influence of independent variables on dependent variables can be represented using logarithmic gains:

$$L(X_j, X_k) = \frac{\partial \ln(X_j)}{\partial \ln(X_k)} = L_{jk} \tag{3}$$

Besides the logarithmic gains of metabolite concentration, the flux logarithmic gains can also be defined. In steady state, the flux of a particular dependent variable X_i (i = 1, 2, ..., n) could be represented:

$$V_i^+ = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{ij}} = \beta_i \prod_{j=1}^{n+m} X_j^{h_{ij}} = V_i^- \tag{4}$$

The logarithmic gains of flux about independent variable X_k should be

$$L(V_i^+, X_k) = g_{ik} + \sum_{j=1}^{n} g_{ij} L_{jk} \tag{5}$$

In the same way, the sensitivity of flux about rate constant α_i should be:

$$S(V_i^+, \alpha_i) = 1 + \sum_{j=1}^{n} g_{ij} S_{ji} \tag{6}$$

Figure 9C,D show separately the influence of the rate constants α_i (i = 1, 2, ..., 9) and independent variables X_j (j = 9, 10, ..., 24) on flux. We can see from Figure 9C that among all of the rate constants, α_1, α_2, α_3, α_8, affect the flux the most significantly, which coincides with our sensitivity analysis results in Figure 9A by computing *RR*. Figure 9D shows that X_9, X_{10}, X_{11}, X_{15}, X_{20}, X_{24} among all of the independent variables influence flux significantly, which is also consistent with the results in Figure 9B.

We also did the correlation analysis between these results of these two methods (flux and robustness ratio) as showed in Figure 10. The results show that there are better correlations between Δ*RR* and flux, and the correlation coefficient reaches to 0.9 and 0.77 (see Supporting Information for more details), respectively, for sensitivity of parameters and logarithmic gains of independent variables when σ = 0.1 (representing variance of the parameters changed).

In Figure 10B,D, one point is off the fit. The reason is that different independent variables could have different influence on the system. By correlation analysis between the results of these two methods (flux and robustness ratio), we found that the trend of the fit is mostly consistent, i.e. when flux increase, the *RR* also increase, which can also be reflected by the correlation coefficient that reaches to 0.77 in Figure 10B. By computing flux and *RR* we obtained similar results of sensitivity analysis, and therefore our approach provides another way to see how rate constants and independent variables influence the stability of system.

Input and Output. We also investigated the relation of input and output for the metabolic pathway. Here we define X_1 as input and X_8 as output. Figure 11A shows input versus output at different diffusion coefficients *D*. From the figure we can see that lower diffusion coefficients correspond to larger *RR* leading to a higher slope of the curve. It also means that the more stable the system is, the larger the slope of the curve is. The input-output curve measures the capability of the response of output to input for the system. The slope represents how sharp the response is. Figure 11B gives clearer results, where we can see that the slope of the input-output curve increases as *RR* increases. In addition, Figure 11C shows the entropy production rate versus the slope of input-output curve or

dx.doi.org/10.1021/sb300020f | *ACS Synth. Biol.* 2012, 1, 229–239

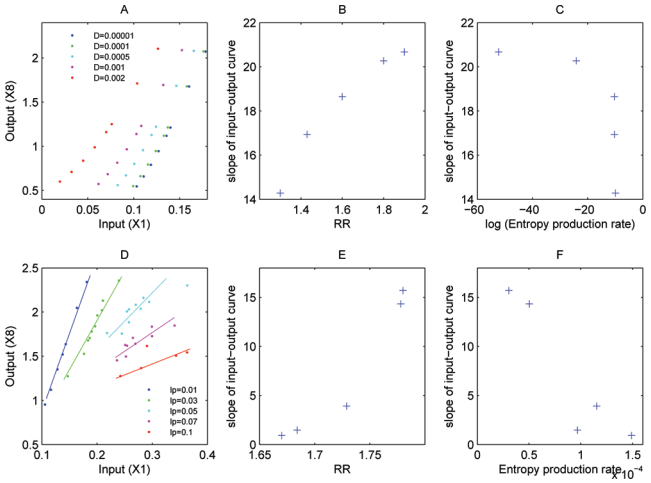


Figure 11. Input versus output for (A) different diffusion coefficients D and (D) different perturbation level of wiring strengths lp. (B, E) Corresponding RR versus slope of input-output response curve in panels A and D. (C, F) Corresponding entropy production rate versus slope of input-output response curve in panels A and D.

response. We can see that when the entropy production rate is small, the slope is big with sharp response and the system is robust, which is consistent with our aforementioned results and discussions about the entropy production rate. Figure 11D−F shows separately input versus output, RR versus slope of the input-output curve or response, and entropy production rate versus slope of input-output curve or response when the perturbation level lp of rate parameters or connection wiring strength is changed. We can see the same trends as the results for external noise for these three figures. Therefore, for both external noise D and perturbation level lp, we can draw the same conclusion that the slope of the input-output curve or response also can be used to evaluate the robustness of the metabolic network. The larger the slope of the input-output curve or the sharper the response is of the output to input, the more stable the system is.

Conclusion. We explored the global natures of a metabolic network in terms of the potential landscape. To reduce the degrees of freedom of the system, a self-consistent mean field approximation method was developed. We used the experimentally inferred rate parameters to study the system by computing RR (robustness ratio) from the underlying landscape topography, Q (similarity order parameters), rmsd, entropy production rates, input-output response, etc. and uncovered that the network is funneled in the space of protein concentrations toward the ground state under the external noise and internal chemical rate perturbations. Robustness ratio (RR) characterizing the landscape topography provides a way to quantitatively measure the global robustness and stability of the metabolic network, and more stable systems have larger RR value. Results of the entropy production rate imply that nature might evolve such that the network is robust against internal and environmental perturbations and performs specific biological functions with less dissipation cost. By sensitivity analysis of parameters, we uncovered some key network structure factors such as kinetic rates or wirings connecting the protein species nodes, which influence the global natures of the system. We also found that there is a strong correlation between the landscape topography and the input-output response. The more stable and robust the metabolic network is, the sharper the response is. Therefore, a funneled landscape provides an optimal criterion to select the suitable parameter subspace of cellular networks, guarantee the robustness, with less dissipations, and perform specific biological functions, which is helpful for the network design. Our approach is general and can be applied to other protein networks and gene regulatory networks,[30] with one or several global minimum.

METHODS

Our aim is to uncover the potential landscape, so we first study the chemical reaction network involved in the metabolic network. We need to take into account the external statistical fluctuations. The statistical natures of the chemical reactions can be captured by the corresponding diffusion equation, which describes the evolution of the networks probabilistically. The diffusion equation is hard to solve due to its inherent huge dimensions. We therefore use the self-consistent mean field approximation to reduce the dimensionality.[7,31] In this way, we can follow the time evolution and steady state probability of the protein concentrations. The steady state probability is closely associated with the potential landscape, which is our target for globally characterizing the system.

Metabolic Network.[24,25] An approach[32] has been used to build up a detailed quantitative model of carbohydrate degradation and oxalacetate formation in *Aspergillus niger*, under conditions of citric acid accumulation.[24]

Chemical reaction equations are based on the power law formalism. By using this formalism the rate of a given process V_i, is written

$$V_i = \alpha_i \prod_{j=1}^{n+m} X_j^{g_{i,j}} \tag{7}$$

Here X_j (j = 1, ..., n for dependent variables and $j = n + 1$, ..., $n + m$ for independent variables) are variables (enzymes, metabolites, effectors) influencing the rates. α and $g_{i,j}$ are the rate constants and kinetic orders of biochemical kinetics. The rate equations for a biochemical system can then be written in terms of power law function approximations as follows:

$$\frac{dX_i}{dt} = \alpha_i \prod X_j^{g_{i,j}} - \beta_j \prod X_j^{h_{i,j}};\ \ i = 1, ..., n \tag{8}$$

The parameters α_i and $g_{i,j}$ are associated with the rate law for net synthesis of X_i, whereas β_i and $h_{i,j}$ are associated with the rate law for net degradation of X_i. Therefore, the two items of the equations represent synthesis and degradation process separately.

Figure 1 represents the main mechanism of the metabolic system.[24,33] Three main metabolic processes are involved in citric acid accumulation by *Aspergillus niger*: (i) breakdown of carbohydrates by the glycolytic pathway to produce pyruvate; (ii) anaplerotic formation of oxalacetate from pyruvate; and (iii) accumulation of citric acid within the tricarboxylic acid cycle. Process (i) and (ii) have been shown to be of great importance both for the final yields and in the metabolic control of the whole process.[33] Here step (i) corresponds to the glucose-6-phosphate (X_1) → fructose-6-phosphate (X_2) → phosphoenol pyruvate (X_3) → pyruvate (X_4) process, and step (ii) corresponds to the pyruvate (X_4) → oxalacetate (X_5) process.

The fundamental equations that characterize the dependent variables of concentrations are given by the mass balance equations. After the kinetic orders and the rate constants are

dx.doi.org/10.1021/sb300020f | ACS Synth. Biol. 2012, 1, 229−239

determined, the model can be represented by the following differential equations:

$$\frac{dX_1}{dt} = 0.459X_8^{0.23}X_9^{0.0009}X_{10}^{1} - 1.833 \times 10^{-3}X_1^{2.194}X_2^{-1.333} X_{11}^{0.2}X_{12}^{0.8}$$

$$\frac{dX_2}{dt} = 3.25 \times 10^{-4}X_1^{2.665}X_2^{-1.667}X_{12} - 1.861X_2^{0.9}X_8^{0.056}X_{13} X_{21}^{0.65}X_{22}^{-0.3}$$

$$\frac{dX_3}{dt} = 3.722X_2^{0.9}X_8^{0.056}X_{13}X_{21}^{0.65}X_{22}^{-0.3} - 0.42X_3^{0.9}X_7^{-0.028} X_8^{-3.196}X_{14}$$

$$\frac{dX_4}{dt} = 0.42X_3^{0.9}X_7^{-0.028}X_8^{-3.196}X_{14} - 2.538X_4^{0.836}X_8^{0.161}X_{15}^{0.509} X_{20}^{0.486}$$

$$\frac{dX_5}{dt} = 1.895X_4^{0.7}X_8^{0.315}X_{15} - 4.23 \times 10^{14}X_5^{3.2}X_6^{-2.2}X_7^{3.788} X_{16}^{0.945}X_{17}^{0.054}$$

$$\frac{dX_6}{dt} = 2.04 \times 10^{15}X_5^{0.7}X_6^{-2.33}X_7^{4.005}X_{16} - 0.005X_6^{0.813}X_{18}^{0.979} X_{19}^{0.02}$$

$$\frac{dX_7}{dt} = 0.204X_3^{0.873}X_7^{-0.028}X_8^{-4.06}X_{14} - 868.709X_5^{1.618} X_6^{-1.136}X_7^{2.203}X_{16}^{0.468}X_{23}^{0.513}$$

$$\frac{dX_8}{dt} = 0.841X_3^{0.9}X_7^{-0.028}X_8^{-3.196}X_{14} - 0.308X_2^{0.225}X_4^{0.179} X_8^{0.257}X_9^{0.294}X_{10}^{0.312}X_{13}^{0.25}X_{15}^{0.256}X_{21}^{0.162}X_{22}^{-0.075}X_{24}^{0.18} \quad (9)$$

Here X_1, ..., X_8 represent eight different variables of concentrations as follows: X_1 (glucose-6-phosphate), X_2 (fructose-6-phosphate), X_3 (phosphoenol pyruvate), X_4 (pyruvate), X_5 (oxaloacetate), X_6 (malate), X_7 (NADH), and X_8 (ATP).

Mean Field Self-Consistent Approximation. The Fokker–Planck or diffusion equation is the one for the time evolution of the probability of some specific state P: $P(X_1, X_2, ..., X_n,t)$, where X_1, X_2, ..., X_n represent the concentration of proteins. The dimensionality of P is exponential: N^M, where N is the segments of each protein concentrations and M is the number of species of proteins. Therefore the equation is not feasible to solve exactly due to large exponential computational dimensionality. We developed a self-consistent mean field approach,[7,12,31] to split the probability into the products of individual ones. First, let us assume $P(X1, X2, ..., Xn,t) = \Pi_i^n P(X_i,t)$, that is, separating the joint distribution $P(x_1, x_2, ..., x_M)$ into the product of each individual one $P(x_1)$, $P(x_2)$, ..., $P(x_M)$ and then self-consistently solving each $P(x_i)$ under the average or mean field influence of others. This effectively reduces the dimensionality from N^M to $N \times M$, and therefore the problem is computationally tractable.

Gaussian Approximation. Self-consistent mean field approximation can reduce the dimensionality of system. In practice, we can further simplify the computations by considering the moment equations. In principle, once we know all moments, then we can construct probability distribution. In many cases, we cannot get all moments. We can start from moment equations and then simply assume specific probability distribution based on physical argument, which means we give some specific relations between moments.[31] For example, Poisson distribution has only one parameter, so we can calculate all other moments from the first moment, mean. Here we use Gaussian distribution as an approximation, and then we need two moments, mean and variance.

For a one-dimensional FPE (Fokker–Planck equation):[21,34]

$$\frac{\partial P(x,t)}{\partial t} = -\frac{\partial}{\partial x}[C(x)P(x,t)] + D\frac{\partial^2}{\partial x^2}[d(x)P(x,t)] \quad (10)$$

Here $C(x)$, $D(x)$ is "drift and diffusion part". For this equation, in weak noise $D \ll 1$, if $D\sigma(t)$ is not large enough to $1/D$ level, the equation for the mean $\langle \dot{x} \rangle$ of the variables x and variance of corresponding variables σ can be written approximately as (through multiplying both sides of the Fokker–Planck Equation by x and x^2 and performing the integral)[21,34]

$$\dot{\mathbf{x}}(t) = C[\mathbf{x}(\mathbf{t})]$$
$$\dot{\sigma}(t) = \sigma(t)\mathbf{A}^{\mathbf{T}}(t) + \mathbf{A}(\mathbf{t})\sigma(\mathbf{t}) + 2\mathbf{D}[\mathbf{x}(\mathbf{t})] \quad (11)$$

Here, $\mathbf{x}$, $\sigma(t)$, and $\mathbf{A}(t)$ are vectors and tensors, and $\mathbf{A}^{\mathbf{T}}(t)$ is the transpose of $\mathbf{A}(t)$. The matrix elements of A is $A_{ij} = (\partial C_i[X(t)])/(\partial x_j(t))$. According to this equations, we can solve $\mathbf{x}(t)$ and $\sigma(t)$. We consider here only diagonal element of $\sigma(t)$ from mean field splitting approximation. Therefore, the evolution of distribution for one variable can be obtained using the mean and variance by Gaussian approximation:

$$P(x,t) = \frac{1}{\sqrt{2\pi}\sigma(t)}\exp -\frac{[x - \bar{x}(t)]^2}{2\sigma(t)} \quad (12)$$

Here, for metabolic network, we first compute the results when the diffusion coefficient is constant, i.e., the diffusion coefficient matrix $D[x(t)]$ is constant diagonal matrix. This condition more likely describes the external environmental fluctuations. In the Supporting Information, we also give the results when the diffusion coefficient is concentration-dependent. In eq 11, replacing $D[x(t)]$ with $d{\cdot}x$ (here d is constant), we can get the concentration dependent diffusion coefficient, which quantifies the influence of intrinsic noise.

The probability obtained above corresponds to one fixed point or basin of attraction. One solution of the equations determines one of the fixed 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, and with different variations. Thus, the total probability is the weighted sum of all these probability distributions ($P(x,t) = w_1P_1(x,t) + w_2P_2(x,t)$, $w_1 + w_2 = 1$). The weighting factors (w_1,w_2) 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.

Finally, once we have the total probability, we can construct the generalized potential landscape by the relationship with the steady state probability $P_{ss}(x)$: $U(x) = -\ln P_{ss}(x)$. 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 function from the steady state probability. In the metabolic

dx.doi.org/10.1021/sb300020f | ACS Synth. Biol. 2012, 1, 229–239

system, every chemical parameter, such as the protein production/decay rates, contributes to the fluctuations of the system. All of these effects are encoded in the total probability distribution, and consequently in the underlying potential landscape.

Entropy Production Rate. In a nonequilibrium open system, there are constant exchanges in energy and information which result dissipations. The dissipation of energy, closely related to the entropy production rate in the steady state, is a global physical characterization of the nonequilibrium system. The entropy formula for the system is well-known:[28]

$$S = -k_B \int P(\mathbf{x}, t) \ln P(\mathbf{x}, t) dx \tag{13}$$

By differentiating the above equation, the increase of the entropy at constant temperature T can be obtained as follows:

$$\begin{aligned} T\dot{S} &= k_B T \int (\ln P + 1) \nabla \cdot \mathbf{J} dx \\ &= -\int (k_B T \nabla \ln P - \mathbf{F}) \cdot \mathbf{J} dx - \int \mathbf{F} \cdot \mathbf{J} dx \end{aligned} \tag{14}$$

here $-\int (k_B T \nabla \ln P - \mathbf{F}) \cdot \mathbf{J}\, dx = e_p$ is the entropy production rate,[28] and $\int \mathbf{F} \cdot \mathbf{J}\, dx = h_d$ is the mean rate of the heat dissipation. $\mathbf{J}(\mathbf{x},t) = \mathbf{F}P - \mathbf{D}\nabla P$ is the probability flux, and $\mathbf{D}$ is diffusion coefficient matrix.[12] In steady state, $\dot{S} = 0$, and the entropy production e_p is balanced by the heat dissipation h_d.

■ ASSOCIATED CONTENT

Supporting Information

This material is free available frree of charge via the Internet at http://pubs.acs.org/.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: ekwang@ciac.jl.cn; jin.wang.1@stonybrook.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

C.H.L. and E.K.W. are supported by the National Natural Science Foundation of China (Grant 21190040) and the 973 project 2009CB930100 and 2010CB933600. J.W. acknowledges support from the National Science Foundation and National Natural Science Foundation of China (Grant 11174105).

■ REFERENCES

(1) Hatzimanikatis, V., Li, C., Ionita, J., and Broadbelt, L. (2004) Metabolic networks: enzyme function and metabolite structure. *Curr. Opin. Struct. Biol. 14*, 300−306.

(2) Guimera, R., and Amaral, L. A. N. (2005) Functional cartography of complex metabolic networks. *Nature 433*, 895−900.

(3) Ma, H., and Zeng, A.-P. (2003) Reconstruction of metabolic networks from genome data and analysis of their global structure for various organisms. *Bioinformatics 19*, 270−277.

(4) Bennett, M. R., Pang, W. L., Ostroff, N. A., Baumgartner, B. L., Nayak, S., Tsimring, L. S., and Hasty, J. (2008) Metabolic gene regulation in a dynamically changing environment. *Nature 454*, 1119−1122.

(5) Zhu, Q., Qin, T., Jiang, Y., Ji, C., Kong, D., Ma, B., and Zhang, H. (2011) Chemical basis of metabolic network organization. *PLoS Comput. Biol. 7*, e1002214.

(6) Heuett, W., Beard, D., and Qian, H. (2008) Linear analysis near a steady-state of biochemical networks: control analysis, correlation metrics and circuit theory. *BMC Syst. Biol. 2*, 44.

(7) Sasai, M., and Wolynes, P. (2003) Stochastic gene expression as a many-body problem. *Proc. Natl. Acad. Sci. U.S.A. 100*, 2374−2379.

(8) Hornos, J., Schultz, D., Innocentini, G., and Wang, J. (2005) Self-regulating gene: an exact solution. *Phys. Rev. E 72*, 051907.

(9) Qian, H., and Reluga, T. C. (2005) Nonequilibrium thermodynamics and nonlinear kinetics in a cellular signaling switch. *Phys. Rev. Lett. 94*, 028101.

(10) Han, B., and Wang, J. (2007) Quantifying robustness of cell cycle network: Funneled energy landscape perspectives. *Biophys. J. 92*, 3755−3763.

(11) Wang, J., Xu, L., and Wang, E. (2008) Potential landscape and flux framework of non-equilibrium networks: Robustness, dissipation and coherence of biochemical oscillations. *Proc. Natl. Acad. Sci. USA. 105*, 12271−12276.

(12) Li, C. H., Wang, J., and Wang, E. (2011) Potential landscape and probabilistic flux of a predator prey network. *PLoS ONE 6*, e17888.

(13) McAdams, H. H., and Arkin, A. (1997) Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. U.S.A. 94*, 814−819.

(14) Elowitz, M. B., and Leibler, S. (2000) A synthetic oscillatory network of transcriptional regulators. *Nature 403*, 335−338.

(15) Swain, P. S., Elowitz, M. B., and Siggia, E. D. (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U.S.A. 99*, 12795−12800.

(16) Thattai, M., and Van, O. A. (2001) Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. U.S.A. 98*, 8614−8619.

(17) Vilar, J. M. G., Guet, C. C., and Leibler, S. (2003) Modeling network dynamics: The lac operon, a case study. *J. Cell Biol. 161*, 471−476.

(18) Paulsson, J. (2004) Summing up the noise in gene networks. *Nature 427*, 415−418.

(19) Hasty, J., Pradines, J., Dolnik, M., and Collins, J. J. (2000) Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci. U.S.A. 97*, 2075−2080.

(20) Hasty, J., Isaacs, F., Dolnik, M., McMillen, D., and Collins, J. J. (2001) Designer gene networks: Towards fundamental cellular control. *Chaos 11*, 207−220.

(21) Van Kampen, N. G. (1992) *Stochastic Processes In Chemistry And Physics*, pp 120−127, North-Holland, Amsterdam.

(22) Li, C. H., Wang, J., and Wang, E. (2011) Landscape and flux decomposition for exploring global natures of non-equilibrium dynamical systems under intrinsic statistical fluctuations. *Chem. Phys. Lett. 505*, 75−80.

(23) Wang, J., Zhang, K., Xu, L., and Wang, E. (2011) Quantifying the waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci. U.S.A. 108*, 8257−8262.

(24) Torres, N. (1994) Modeling approach to control of carbohydrate metabolism during citric acid accumulation by aspergillus niger: I. Model definition and stablility of the steady state. *Biotechnol. Bioeng. 44*, 104−111.

(25) Torres, N. (1994) Modeling approach to control of carbohydrate metabolism during citric acid accumulation by aspergillus niger: II. Sensitivity analysis. *Biotechnol. Bioeng. 44*, 112−118.

(26) Ao, P. (2004) Potential in stochastic differential equations: Novel construction. *J. Phys. A: Math. Gen. 37*, L25−L30.

(27) Kacser, H., and Burns, J. A. (1973) The control of flux. *Symp. Soc. Exp. Biol. 27*, 65−104.

(28) Qian, H. (2001) Mesoscopic nonequilibrium thermodynamics of single macromolecules and dynamic entropy-energy compensation. *Phy. Rev. E 65*, 0161021−0161025.

(29) Torres, N., and Voit, E. (2005) *Pathway Analysis and Optimization in Metabolic Engineering*, Chemical Industry Press: Beijing.

(30) Tyson, J., and Novak, B. (2001) Regulation of the eukaryotic cell cycle: molecular antagonism, hysteresis, and irreversible transitions. *J. Theor. Biol. 210*, 249−263.

(31) Lapidus, S., Han, B., and Wang, J. (2008) Intrinsic noise, dissipation cost, and robustness of cellular networks: The underlying energy landscape of mapk signal transduction. *Proc. Natl. Acad. Sci. U.S.A. 105*, 6039−6044.
(32) Savageau, M. A. (1976) *Biochemical System Analysis: A Study of Function and Design in Molecular Biology*, Addison-Wesley Pub. Co., Boston.
(33) Mattey, M. (1992) The production of organic acids. *CRC Crit. Rev. Biotechnol 12*, 87−132.
(34) Hu, G. (1994) *Stochastic Forces And Nonlinear Systems*, pp 29−74, Shanghai Scientific And Technological Education Press, Shanghai.