Transition Path, Quasi-potential Energy Landscape and Stability of Genetic Switches

Cheng Lv,1 Xiaoguang Li,2 Fangting Li,1,3,† and Tiejun Li2,4,†

1School of Physics, Peking University, Beijing 100871, China
2LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, China
3Center of Quantitative Biology, Peking University, Beijing 100871, China
4Beijing International Center for Mathematical Research, Beijing 100871, China

One of the fundamental cellular processes governed by genetic regulatory networks in cells is the transition among different states under the intrinsic and extrinsic noise. Based on a two-state genetic switching model with positive feedback, we develop a framework to understand the metastability in gene expressions. This framework is comprised of identifying the transition path, reconstructing the global quasi-potential energy landscape, analyzing the uphill and downhill transition paths, etc. It is successfully utilized to investigate the stability of genetic switching models and fluctuation properties in different regimes of gene expression with positive feedback. The quasi-potential energy landscape, which is the rationalized version of Waddington potential, provides a quantitative tool to understand the metastability in more general biological processes with intrinsic noise.

PACS numbers: 87.18.Cf, 02.50.Ey, 82.39.-k, 87.17.Aa

In a cell, the reactions underlying gene expression involve small numbers of molecules, such as DNA, mRNAs, and transcription factors, so the stochasticity in gene regulation process is inevitable even under constant environmental conditions [11, 2]. Recent progresses in single-cell observations and analysis demand new quantitative methods to characterize the regulatory mechanism in both prokaryotes [3] and eukaryotes [4] from the cellular stochasticity.

Previous kinetic studies of cellular stochasticity have been formulated by using the generating function [5], system size expansion [6, 7], large deviation theory [8, 9], or by employing WKB approximation to the chemical master equations (CMEs) [10–12], etc. However, only few of them take account of transcriptional noise explicitly. Some recent studies have shown that correlations between mRNA and protein levels do not always perform equally well in revealing regulatory relationships [13] and the consideration of mRNA has intensive effect on the switching times [15, 16]. Since Waddington's “epigenetic landscape” proposed in 1957 [17], the energy landscape have been widely used to provide pictorial illustration of the dynamics and evolution of genetic regulatory systems [2, 8, 18]. Thus it is important and natural to develop a general methodology which can effectively determine the key features of a gene expression system, such as constructing the corresponding “Waddington potential”, identifying the transition paths between metastable states and computing the transition rates, etc.

In this letter, we present a framework to understand the metastability of the genetic switches in gene expression based on the large deviation theory for Markov processes [19, 21]. By explicitly taking into account the mRNA noise, we obtain the most probable transition paths for off-to-on and on-to-off genetic switches through the geometric minimum action method (gMAM) [22]. Furthermore, we reconstruct the global quasi-potential energy landscape, which is the rationalized version of the Waddington potential, via the computation of the local quasi-potentials. We analyze the properties of transition paths, discuss their relation to the quasi-potential energy landscape and the deterministic dynamics, and compare the differences between our approach and others in previous literatures. Our analytical results agree well with the Monte Carlo simulations. We also discuss the choice of the system size and its finite effect via the transition path theory (TPT) [23]. From the authors’ opinion, this framework is generally applicable for studying transitions between stable-saddle-stable fixed points with jump type noise generated by Gillespie’s birth-death dynamics [24]. It is successfully utilized to investigate the stability of genetic switching models and fluctuation properties in different regimes of gene expression with positive feedback, which leads to interesting biological insights.

Let us focus on a two-state gene expression model in Fig. 1(a) following the previous study [12], where the transitions of active and inactive promoter states are controlled by the protein via a positive feedback. The transition rates between active and inactive states are \( k_{on}(n) \equiv f(n) \) and \( k_{off}(n) \equiv g(n) \), where \( n \) is the protein copy number, \( on \) and \( off \) denote the high and low states of protein and mRNA respectively. Here the transition rates are in the form of Hill-function as

\[
\begin{align*}
    f(n) &= k_{0} \min + (k_{\max}^0 - k_{\min}^0) n^h_1/(n^{h_1}_50 + n^{h_1}), \\
    g(n) &= k_{\max} - (k_{\max} - k_{\min}) n^h_2/(n^{h_2}_50 + n^{h_2}),
\end{align*}
\]

where \( n_{50} \) is the curve’s midpoint, and \( h_1 = h_2 = h \) are chosen for simplicity.

A deterministic description of the genetic switching model in Fig. 1 is given by the equations

\[
\begin{align*}
    \dot{M} &= af(N)/[f(N) + g(N)] - \gamma M, \\
    \dot{N} &= \gamma b M - N,
\end{align*}
\]

through quasi-steady state approximation, where \( M, N \) denote the mean number of mRNAs and proteins, re-
The large deviation theory gives a good quantitative description of rare events \[19, 22, 25\]. The most probable transition path is given by minimizing an action functional characterized by a Lagrangian \( L \). For the Gillespie’s birth-death dynamics, \( L \) has no closed form and only its dual Hamiltonian can be given as

\[
H(x,p) = \sum_{j=1}^{N} a_j(x)(e^{\nu_j} - 1),
\]

where \( a_j \) is the rate (or propensity) and \( \nu_j \) is the state-change (or stoichiometric) vector, \( j = 1, 2, \ldots, N \). However, this description has difficulty when taking DNA into consideration, for normally there are too few DNA copies in a living cell that the straightforward application of the above formulation is not valid.

Now we adopt Assaf et al’s approach \[12\] by considering the stationary distribution where \( \tilde{P}_{m,n} = \tilde{Q}_{m,n} = 0 \) and eliminating \( Q_{m,n} \) to obtain

\[
0 = \{ \mathbf{A} + g(n)^{-1}[\mathbf{A} + a(\mathbf{E}_m^{-1} - 1)](f(n) - \mathbf{A}) \} \tilde{P}_{m,n}. \tag{4}
\]

Define the concentration variable \( x = m/K, y = n/K \) and the quasi-potential \( S(x,y) \) \[20\], we plug WKB ansatz

\[
\tilde{P}_{m,n} \equiv P(x,y) \sim \exp[-KS(x,y)]
\]

into Eq. \( 4 \) to get a stationary Hamilton-Jacobi equation \( H(x,y,\partial_xS,\partial_yS) = 0 \) to the leading order with the Hamiltonian

\[
H = A + \tilde{g}(y)^{-1}[A + b^{-1}(e^{\nu_y} - 1)](f(y) - A),
\]

where the function \( A = A(x,y,p_x,p_y) = y(e^{-\nu_y} - 1) + \gamma x(e^{-\nu_y} - 1) + \gamma bx(e^{\nu_y} - 1) \) and the feedback functions are now rescaled as \( f(y) = f(y)/K, \tilde{g}(y) = g(y)/K \). The classical Hamilton-Jacobi theory enables one to solve the quasi-potential \( S(x,y) \) with different formulations such as the variational methods, integrating Hamiltonian dynamics or directly solving PDEs, in which we have the connection \( p_x = \partial_xS, p_y = \partial_yS \) between the momenta and the quasi-potential \[26\]. It is worth asking whether the choice of the large parameter \( K \) affects the final results since any choice is artificial in practice. An affirmative answer is given in SM that only the scaling matters and the final systems are equivalent with respect to different choices of the large parameter \( K \).

Now we use the powerful gMAM algorithm \[22\] to compute the quasi-potential by minimizing the action functional with the obtained Hamiltonian \[6\] (some details about the gMAM can be referred to SM). We choose the two stable points solved by Eqs. \[1\] as our starting and ending points, then compute the switching paths starting from either of the two states (see Fig. \[2\]). These switching paths predict well with the switching trajectories obtained from the MC simulations.

Figure \[2\] shows clearly that when switch occurs, the switching trajectory prefers to be around the most probable path characterized by the Hamiltonian \[4\]. Furthermore, the off-to-on and on-to-off paths are not identical, suggesting that the switching paths are irreversible.
which is different from the original Waddington picture [17]. The irreversibility is fundamental in chemical reaction kinetics due to the non-gradient nature of the dynamical system. Figure 2 also tells that both switching paths pass through the same bottleneck, i.e. the saddle point given by Eqs. (1). To further characterize the switching path, we note that the transition path is also given by the Hamilton’s equations $\dot{x} = \nabla_p H$, $\dot{p} = -\nabla_x H$, where $x = (x, y)$, $p = (p_x, p_y)$. Based on the fact $H(x, y, 0, 0) \equiv 0$, we obtain $\nabla_x H \equiv 0$ when $p = 0$. At the saddle point in any transition path, we have $p = 0$ [22], and thus $p = 0$ along the whole downhill path. With this result we obtain the downhill equations $\dot{x} = \nabla_p H(x, 0)$, which exactly corresponds to the deterministic dynamics [1] with a renormalization factor $\tilde{\phi} = \phi(x)/[\tilde{f}(x) + \tilde{\phi}(y)]$ of time since this only involves the DNA inactive state. This fact explains that after climbing on the downhill path is then given by $\dot{x} = \nabla_p H(x, \nabla_x S)$ which is in general different from the downhill path unless the system is of gradient type.

Besides giving the minimum action path, gMAM also contributes the action value at each point in the path, which enables us to calculate the mean switching time (MST) $\tau$ from either stable state. Denote the quasi-potential energy barrier $\Delta S_{on} = S_0 - S_{on}$, where $S_0$ is the action at the saddle point and $S_{on}$ is the action at the on state. Then the MST $\tau_{on}$ from on-to-off transition can be roughly estimated from an asymptotic analysis

$$\tau_{on} \approx T_{on} e^{K \Delta S_{on}},$$

where $T_{on}$ is a prefactor which is independent of $K$ in many cases. Although for one dimensional systems the prefactor of MST can be obtained [27], there are no available results in high dimensions because of the geometry problem in more than one spatial dimension and the non-gradient nature of the system [28,29]. Therefore we compare the MC simulations with the exponential time part and adjust the prefactor $T_{on}$ to fit the numerical results. Figure 3(a) and (b) demonstrate the MST versus the variation of parameters $b$ and $n_{50}$, respectively. It shows that the MST is excellently predicted by Eq. (7) up to a slowly varying prefactor. The analysis for $\tau_{off}$ is similar. Our results in Fig. 3(a) show that the MST from both off-to-on and on-to-off states decrease rapidly when translation rate $b$ is increased. And when $n_{50}$ increases in the considered interval, MST from off to on state increases exponentially, while MST from on to off state decreases exponentially. When $n_{50}$ is about 1004, these two MSTs are equal. These results may provide hints to the range of kinetic reaction rates.

Based on the obtained quasi-potential values for each point from on and off states as starting point, we can reconstruct the global quasi-potential energy landscape for genetic switching model. The two and three dimensional view of the quasi-potential energy landscape is shown in Fig. 4 and the reconstruction details can be found in SM. The reconstruction for more general stochastic dynamical systems can be referred to [20,30]. In Fig. 4 we observe that the on and off states correspond to two local minimum on the quasi-potential energy landscape, the saddle of the deterministic dynamical system exactly corresponds to the saddle point on the quasi-potential energy landscape, too. As discussed in [20], we can obtain that the most probable uphill path satisfies $\phi = a(\psi) \cdot \nabla U(\phi) + l(\phi)$, and the most probable downhill path satisfies $\psi = b(\psi) = -a(\psi) \cdot \nabla U(\psi) + l(\psi)$ if the underlying stochastic dynamics has the form $\dot{x} =$
order thus get the Gaussian approximation
\[ P(x, y) \simeq \frac{1}{(2\pi)^{1/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (x - \mu)^{\top} \Sigma^{-1} (x - \mu) \right\}. \quad (8) \]

Here, \( x = (x, y), \mu = (\mu_x, \mu_y), \Sigma = (\Sigma_{ij})_{2 \times 2}, \) and \( |\Sigma| \) is the determinant of matrix \( \Sigma \). Eq. (8) holds only in the vicinity of the on state with standard deviations \( \sigma_m = (KS_{yy}/|\Sigma|)^{1/2}, \sigma_n = (KS_{xx}/|\Sigma|)^{1/2} \). With the \( \sigma_n \) and \( \sigma_m \) above, we can easily obtain the SMR as shown in Fig. 5.

We distinguish among several mechanisms of DNA/promoter activation in genetic switches by selecting different promoter transition rates \( a \) and mRNA degradation rates \( \gamma \), which is similar to the previous work of Raser and O’Shea [4]. The above model in Fig. 1 is set with the parameters \( a=106.7 \) and \( \gamma = 2.0 \). We set a slow chromatin-remodeling eukaryotic model with \( k_{0}^{\text{max}} = k_{1}^{\text{max}} = a/10, k_{0}^{\text{min}} = k_{1}^{\text{min}} = a/1000, \) and \( \gamma = 1.5 \), while a prokaryotic model is set as \( k_{0}^{\text{max}} = k_{1}^{\text{max}} = 10a, k_{0}^{\text{min}} = k_{1}^{\text{min}} = a/10, \) with a higher mRNA degradation rate \( \gamma = 15 \). The results in Fig. 5 show that the noise level in eukaryotic model is almost always bigger than that in prokaryotic model. And the relationship between SMR and mean expression level is different from the no-feedback models [4], especially for the changing of transcription rate \( a \). It is interesting to find that when increasing the transcription rate \( a \), noise level of on state decreases in the model with positive feedback, while protein noise increases in one-state chromatin-remodeling eukaryotic model without positive feedback (Case I in Raser and O’Shea’s work). In SM, we also compare the SMR curves of different promoter transition rates \( a \) with fixed \( \gamma = 2.0 \) (Fig. S4) and provide the stochastic features of mRNA (Fig. S3).

Our methodology above is based on large deviation theory, which requires that the system size tends to infinity. For more general and realistic cases, the system size is finite, thus the transition path theory (TPT) can
be applied to find the most probable transition paths \cite{23}. Here we choose the nearest 4 lattice points around starting and ending point in gMAM as our starting and ending sets and perform the TPT computations (see Fig. 6). The implementation details can be found in SM. Fig. 6 shows that TPT result fits MC simulation slightly better than gMAM. TPT result in Fig. 6(a) also provides interesting details to understand the uphill trajectory from off to on state. In the beginning the off-to-on switch, the system is mostly at the active DNA/promoter state thus the mRNA level increases before the translation of protein. While in the beginning of on-to-off switch in Fig. 6(b), we observe opposite mechanism with inactive DNA/promoter and first decrease of mRNA. This explains why the uphill transition path from off to on state is convex and it is concave from on state. In the beginning the system is in DNA active state while blue line means DNA is inactive. The parameter setup is the same as in Fig. 4.

In this letter, we have presented an analytical framework to reconstruct the quasi-potential energy landscape of genetic switching system while explicitly taking mRNA noise into account. We further analyze the properties of the transition paths and clarify the relation with the gradient-curl decomposition in the previous literatures. This global potential, which is a rationalized version of Waddington potential, provides a quantitative tool to understand the metastability in more general biological processes with intrinsic noise. The applications to other biological systems such as complex cellular decision making process and the development process of cells will be investigated in the future.

The authors are grateful to Weinan E, Xiang Zhou, Qi Ouyang and Hongli Wang for helpful discussions. The work is supported by NSFC grants no.11174011, 11021463 (F.Li), 11171009 and 91130005 (T.Li).

\begin{thebibliography}{99}
\bibitem[1]{1} B. Munsky, G. Neuert, and A. van Oudenaarden, \textit{Science} \textbf{336}, 183 (2012).
\bibitem[2]{2} G. Balázs, A. van Oudenaarden, and J.J. Collins, \textit{Cell} \textbf{144}, 910 (2011).
\bibitem[3]{3} M.B. Elowitz \textit{et al.}, \textit{Science} \textbf{297}, 1183 (2002).
\bibitem[4]{4} J.M. Raser and E.K. O’Shea, \textit{Science} \textbf{304}, 1811 (2004).
\bibitem[5]{5} M. Thattai and A. van Oudenaarden, \textit{Proc. Natl. Acad. Sci. USA} \textbf{98}, 8614 (2001); T.B. Kepler and T.C. Elston, \textit{Biophys. J.} \textbf{81}, 3116 (2001).
\bibitem[6]{6} J. Paulsson, \textit{Nature} \textbf{427}, 415 (2004).
\bibitem[7]{7} N.G. van Kampen, \textit{Stochastic Processes in Physics and Chemistry}, North-Holland, Amsterdam, New York and Oxford, 1981.
\bibitem[8]{8} J. Wang, K. Zhang and E. Wang, \textit{J. Chem. Phys.} \textbf{133}, 125103 (2010).
\bibitem[9]{9} D.M. Ronna \textit{et al.}, \textit{Phys. Rev. E} \textbf{71}, 011902 (2005); D. Liu, \textit{J. Comp. Phys.} \textbf{227}, 8672 (2008); C.C. Govern, M. Yang, and A.K. Chakraborty, \textit{Phys. Rev. Lett.} \textbf{108}, 058102 (2012).
\bibitem[10]{10} A.M. Walczak, M. Sasai, and P.G. Wolynes, \textit{Biophys. J.} \textbf{88}, 828 (2005).
\bibitem[11]{11} J.E.M. Horinos \textit{et al.}, \textit{Phys. Rev. E} \textbf{72}, 051907 (2005).
\bibitem[12]{12} M. Assaf, E. Roberts, and Z. Luthey-Schulten, \textit{Phys. Rev. Lett.} \textbf{106}, 248102 (2011).
\bibitem[13]{13} S.J. Gandhi \textit{et al.}, \textit{Nat. Struct. Mol. Biol.} \textbf{18}, 27 (2011).
\bibitem[14]{14} Y. Taniguchi \textit{et al.}, \textit{Science} \textbf{329}, 533 (2010).
\bibitem[15]{15} P. Mehta, R. Mukhopadhyay, and N.S. Wingreen, \textit{Phys. Biol.} \textbf{5}, 026005 (2008).
\bibitem[16]{16} C. Zong \textit{et al.}, \textit{Mol. Syst. Biol.} \textbf{6}, 440 (2010).
\bibitem[17]{17} C. H. Waddington and H. Kacser, \textit{Classical mechanics}, 2nd edition, Addison-Wesley, 1980.
\bibitem[18]{18} H. Goldstein, \textit{Classical mechanics}, 2nd edition, Addison-Wesley, 1980.
\bibitem[19]{19} C.R. Doering \textit{et al.}, \textit{J. Phys.: Cond. Mat.} \textbf{19}, 065145 (2007); B. Meerson and P.V. Sasorov, \textit{Phys. Rev. E} \textbf{78}, 060103(R) (2008); C. Escudero and A. Kamenev, \textit{Phys. Rev. E} \textbf{79}, 041149 (2009).
\bibitem[20]{20} T. Naeh, \textit{et al.}, \textit{SIAM J. Appl. Math.} \textbf{50}, 595 (1990).
\end{thebibliography}