

RESEARCH ARTICLE

Open Access

Distinguishing the rates of gene activation from phenotypic variations



Ye Chen¹, Cheng Lv², Fangting Li^{2,3} and Tiejun Li^{1*}

Abstract

Background: Stochastic genetic switching driven by intrinsic noise is an important process in gene expression. When the rates of gene activation/inactivation are relatively slow, fast, or medium compared with the synthesis/degradation rates of mRNAs and proteins, the variability of protein and mRNA levels may exhibit very different dynamical patterns. It is desirable to provide a systematic approach to identify their key dynamical features in different regimes, aiming at distinguishing which regime a considered gene regulatory network is in from their phenotypic variations.

Results: We studied a gene expression model with positive feedbacks when genetic switching rates vary over a wide range. With the goal of providing a method to distinguish the regime of the switching rates, we first focus on understanding the essential dynamics of gene expression system in different cases. In the regime of slow switching rates, we found that the effective dynamics can be reduced to independent evolutions on two separate layers corresponding to gene activation and inactivation states, and the transitions between two layers are rare events, after which the system goes mainly along deterministic ODE trajectories on a particular layer to reach new steady states. The energy landscape in this regime can be well approximated by using Gaussian mixture model. In the regime of intermediate switching rates, we analyzed the mean switching time to investigate the stability of the system in different parameter ranges. We also discussed the case of fast switching rates from the viewpoint of transition state theory. Based on the obtained results, we made a proposal to distinguish these three regimes in a simulation experiment. We identified the intermediate regime from the fact that the strength of cellular memory is lower than the other two cases, and the fast and slow regimes can be distinguished by their different perturbation-response behavior with respect to the switching rates perturbations.

Conclusions: We proposed a simulation experiment to distinguish the slow, intermediate and fast regimes, which is the main point of our paper. In order to achieve this goal, we systematically studied the essential dynamics of gene expression system when the switching rates are in different regimes. Our theoretical understanding provides new insights on the gene expression experiments.

Keywords: Gene expression, Positive feedback, Energy landscape, Flow cytometry experiment

Background

Stochasticity is an inherent property of living cells. Stochastic perturbation comes not only from the noisy environment that cells live in, but also from all processes going on inside cells [1]. Especially when the copy number of molecules like mRNA or protein gets smaller, stochastic fluctuations can have significant effect on the

behavior of a cell. There have been many studies focused on discovering the effect of noise in living systems [2, 3], and some of the results show that cells do not only evolve to resist noise perturbation better, but also form massive mechanisms which are the benefits of stochasticity [4, 5]. Genetic switching is a typical example of noise-driven cellular behavior which can be performed spontaneously under particular conditions in living cells, such as the switch of *lac* operon in *Escherichia coli* [6], the formation of biofilm [7], and the entrance into competence state [8] in *Bacillus subtilis*, etc.

*Correspondence: tieli@pku.edu.cn

¹LMAM and School of Mathematical Sciences, Peking University, No. 5 Yiheyuan Road, 100871 Beijing, China

Full list of author information is available at the end of the article

Several recent studies discovered strong connections between noise and gene regulation mechanisms through measuring variability in protein and mRNA levels [9, 10]. Both experimental and theoretical results suggested that the switching rates of gene between its active and inactive states have significant effects on the noise of gene expression [11, 12] and are highly responsible to different gene expression dynamics. Previous theoretical studies mainly focused on the regime with fast switching rates [13–15], however, recently some unique phenomena are observed especially when switching rates are slow, either due to the chromatin remodeling process [16] or the transportation of transcription factor into nucleus [17]. Thus, it is important and desirable to provide an approach which can effectively identify the key dynamical features of gene expression system when switching rates are slow, and to distinguish the regime of gene regulatory networks from different phenotypic variations.

In this paper, we aim at providing a method to distinguish the rates of genetic switches from phenotypic variations. First, we systematically investigated the dynamical and stochastic properties of different genetic regulatory models in a comparative way when biochemical rates for gene activation/inactivation, transcription, translation, and mRNA/protein degradation vary over a wide range. We focused on a two-state gene-expression model with positive feedback loop, which is simplified from a four-state gene-expression network taking account of chromatin opening/closing step and regulatory protein binding/unbinding step [16, 18]. In the regime when the rates of gene activation/inactivation are relatively small compared with the synthesis/degradation rates of mRNAs and proteins, we found that the effective dynamics can be reduced to independent evolutions on two separate layers. We reconstructed the energy landscape of the system through Gaussian mixture approximations, and it is qualitatively different from the previously known picture in [15, 19]. In the regime with relatively fast switching rates, we briefly discussed about the system based on our previous study with large deviation theory [15]. In the regime with intermediate switching rates, we utilized the mean switching time (MST) to investigate the stability of the system in different parameter ranges. In the following paragraphs, we will call these three regimes as slow, fast, and intermediate regimes for convenience. Finally, based on our theoretical understanding of the essential dynamics of the slow, intermediate and fast regimes, we designed a simulation experiment to distinguish the three regimes. We identified the intermediate regime from the fact that cellular memory is much weaker in this regime than the other two regimes. And we distinguished the slow and fast regimes through their different response behavior with respect to switching rates perturbations. Our study gives

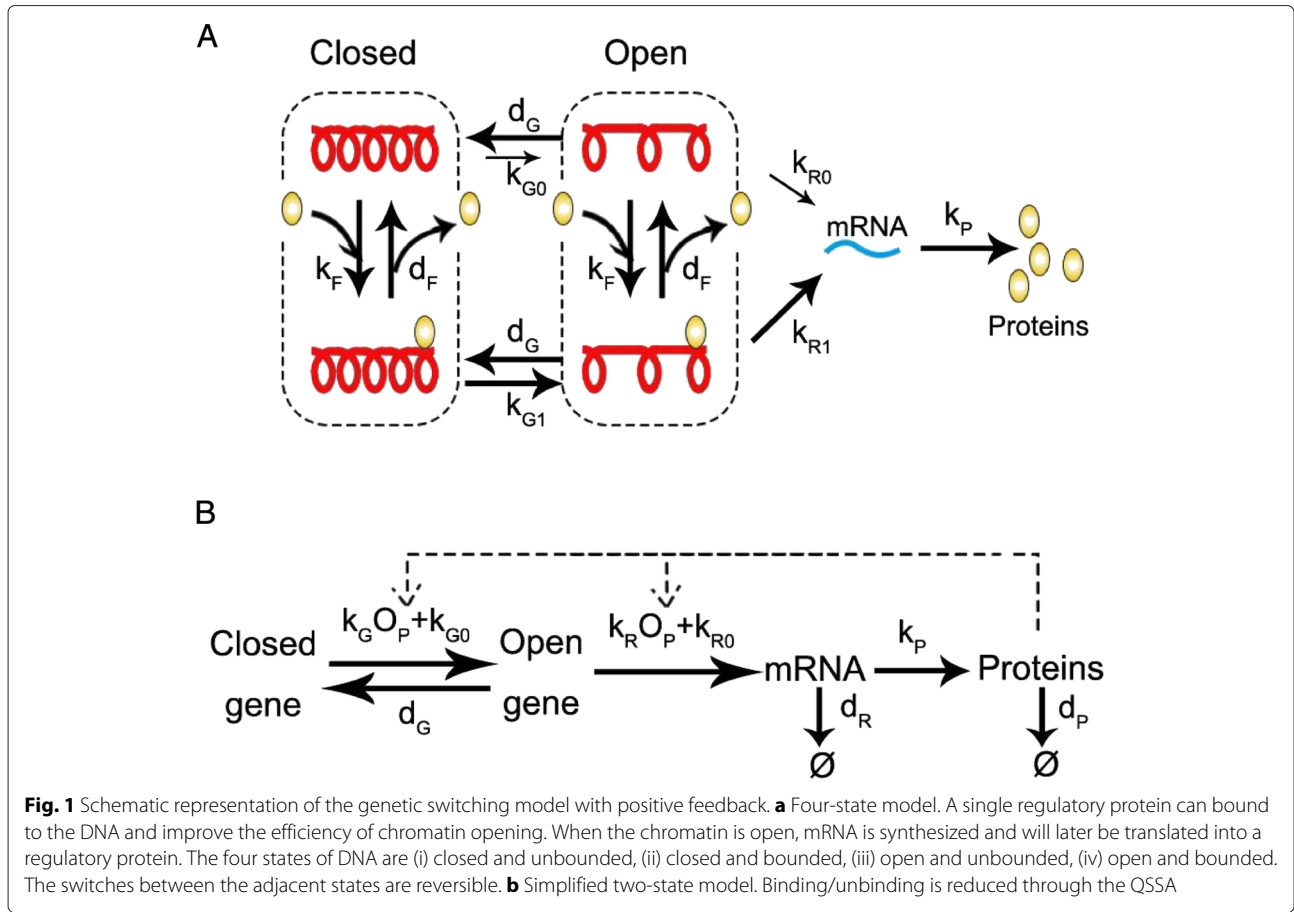
a global characterization of the genetic switching dynamics in various cells and has potential guidance in real experiments.

Methods

We consider a four-state gene-expression model with positive feedback loop (Fig. 1a). Proteins are self-activators in this system, since they can bind to DNA and activate the transcription step. Chromatin opening becomes relatively easy when DNA is bounded by a monomer regulatory protein. Open chromatin structure is permissible for transcription whereas closed chromatin is not. The rate of mRNA production reaches its maximum on the bounded and open state of DNA. Additionally, proteins and mRNAs are subject to degradation, which is not shown in Fig. 1a. Assuming that binding and release of regulatory proteins are in fast equilibrium [20, 21], this four-state model can be simplified into a two-state model (Fig. 1b) through the quasi-steady-state approximation (QSSA) (see Additional file 1 for derivation details) [22, 23]. We will say that the gene is *open* (*closed*) when the chromatin is open (closed), and the system switches between open and closed gene states through gene activation/inactivation steps.

In the two-state model, there are six reactions occurring among four chemical species. All of the reactions involved in this system are summarized in Table 1. Here m and n are the numbers of protein molecules and mRNA molecules, respectively. The jump rate into the open state and the transcription rate are controlled by the relative occupancy of protein at DNA binding site (i.e. O_P), whereas the jump rate into the closed state is not affected by the number of proteins. O_P has the form $O_P = n/(n + K)$, where K is a constant by reduction and represents the strength of positive feedback. It is actually the conditional probability of state (iv) in Fig. 1a given that the chromatin is open. The rates $k_G O_P + k_{G0}$ and $k_R O_P + k_{R0}$ characterize the gene activation and mRNA synthesis, which are dependent of the number of proteins. The constants d_G , k_P , d_R and d_P correspond to the rates of gene inactivation, protein synthesis, mRNA and protein degradation. In fact, if we assume that dimer proteins (or other multimer-type of proteins) regulate the gene instead of the monomer protein, O_P will have the form $O_P = n^k/(n^k + K^k)$. The value of k will not influence the essential dynamics of the gene expression model, on condition that we choose proper range of parameters with respect to different k . Under some reasonable assumptions, most of the derivations and results in the next section are still correct (see Additional file 1 for details).

Due to the intrinsic noise, the dynamics of the system can be described by the chemical master equation (CME) [1, 24–27] as



$$\begin{aligned} \frac{dP_0(m, n)}{dt} = & - \left(k_{G0} + k_G \frac{n}{n+K} \right) P_0(m, n) \\ & + d_G P_1(m, n) + k_P [P_0(m, n-1) - P_0(m, n)] \\ & + d_R [(m+1) P_0(m+1, n) - m P_0(m, n)] \\ & + d_P [(n+1) P_0(m, n+1) - n P_0(m, n)], \end{aligned} \tag{1}$$

$$\begin{aligned} \frac{dP_1(m, n)}{dt} = & \left(k_{G0} + k_G \frac{n}{n+K} \right) P_0(m, n) \\ & - d_G P_1(m, n) + k_P [P_1(m, n-1) - P_1(m, n)] \\ & + \left(k_{R0} + k_R \frac{n}{n+K} \right) [P_1(m-1, n) - P_1(m, n)] \\ & + d_R [(m+1) P_1(m+1, n) - m P_1(m, n)] \\ & + d_P [(n+1) P_1(m, n+1) - n P_1(m, n)], \end{aligned} \tag{2}$$

Table 1 Reactions in the two-state gene expression network with positive feedback

Number	Reaction
1	DNA _{open} $\xrightarrow{d_G}$ DNA _{closed}
2	DNA _{closed} $\xrightarrow{k_G O_P + K_{G0}}$ DNA _{open}
3	DNA _{open} $\xrightarrow{k_R O_P + K_{R0}}$ DNA _{open+m}
4	$m \xrightarrow{k_P} n$
5	$m \xrightarrow{d_R} \emptyset$
6	$n \xrightarrow{d_P} \emptyset$

where $P_\alpha(m, n)$ stands for the probability that there are m mRNA molecules and n protein molecules in the system when the gene is open ($\alpha = 1$) or closed ($\alpha = 0$). One possibility to study the behavior of the system is to numerically solve Eqs. 1-(2) by truncating the domain and setting the boundary conditions $P_\alpha(m, n) = 0$ when $m \geq M, n \geq N$ (M, N is large enough, $\alpha = 0, 1$). But simply obtaining this solution does not mean we understand the essential mechanism of the dynamics, and the computational effort may be huge when the mean number of molecules is quite large in this system. Thus, we aim to find other ways to learn about the dynamics.

We define a key parameter κ to compare two characteristic timescales of the problem: the timescale d_P^{-1} on which the gene state changes and the timescale d_G^{-1} on which the number of protein and mRNA molecules changes, where d_P^{-1} is the lifetime of a single protein and d_G^{-1} is the lifetime of the open gene state. The ratio $\kappa = d_G/d_P$ can take any positive value, which will lead to notable differences in the dynamics. Small ratios ($\kappa \ll 0.01$) describe long timescales of the change in gene states, compared with the timescale on which mRNA and protein numbers change. In this regime, there is enough time for the system to reach a steady state when gene is either open or closed. Transition from one steady state to another always occurs immediately after the genetic switches. Large values of the ratio ($\kappa \gg 1$) indicate that the gene states change very rapidly, whereas the synthesis and degradation of protein and mRNA molecules are relatively slow. As κ goes to infinity, the system exhibits the deterministic features [13, 15]. When κ is neither large nor small enough, genetic switches, transcription and translation will take place on the same timescale. The dynamics is complicate in this regime, and analytical results are hard to reach. So we resort to numerical simulations and extract useful information. In all of the three regimes, metastable states may occur with proper choice of parameters. Indeed, we will utilize a set of biologically relevant parameters in later computations (see Additional file 1 for details). For convenience, we will call the steady state with relatively large number of proteins as *on* state, whereas the steady state with relatively small number of proteins is *off* state.

Results

Case A: slow regime

We first consider the slow regime where the gene states change very slowly compared with the synthesis and degradation of proteins and mRNAs. We start from constructing the energy landscape, which is a useful way to illustrate the dynamics of the system. We simply take the definition of the potential as $U(m, n) = -\ln P(m, n)$, which is commonly used by Wang et al. [14, 28, 29]. Here P is the invariant distribution with the form $P(m, n) = P_0(m, n) + P_1(m, n)$. Solving the CME of the system (1)-(2) with brutal force, we get the energy landscape as shown in Fig. 2a. It contains two attracting basins, among which the one with large number of molecules is related to the open state, whereas the other one located at the origin is related to the closed state. As the global energy landscape reflects the steady state distribution of the system, it can not tell us non-equilibrium information. Below we will address this issue by reducing the dynamics and revealing the essential nature of the system.

A careful observation shows that the gene will stay on a single state for a very long time and then make transition

to another state. So we decouple the landscape into two separated layers according to the gene state (Fig. 2b). The upper layer stands for the open state, whereas the lower layer stands for the closed state. Transitions between two layers do not often occur. Each layer has only one steady state, which can be found through deterministic rate equations

$$\begin{aligned} \frac{dm}{dt} &= \alpha \left(k_{R0} + k_R \frac{n}{n+K} \right) - d_R m, \\ \frac{dn}{dt} &= k_P m - d_P n, \end{aligned} \quad (3)$$

where α takes value 0 if the gene is closed and value 1 if the gene is open. The steady states of Eq. 3 when $\alpha = 0$ and 1 characterize the attraction points in the overall energy landscape and the relaxation from one steady state to the other corresponds to the transition paths. If the noise effect can not be neglected, we can further get the diffusion approximation beyond Eq. 3 (see Additional file 1 for details). When the system size is large, the dynamical process on each level can be approximated by diffusion process [30–32]. We consider the concentration variables $x = m/V$, $y = n/V$ on the lattice \mathbb{N}^2/V , where V is the volume size. Define $\mathbf{x} = (x, y)$, then the first and second moments of the process satisfy

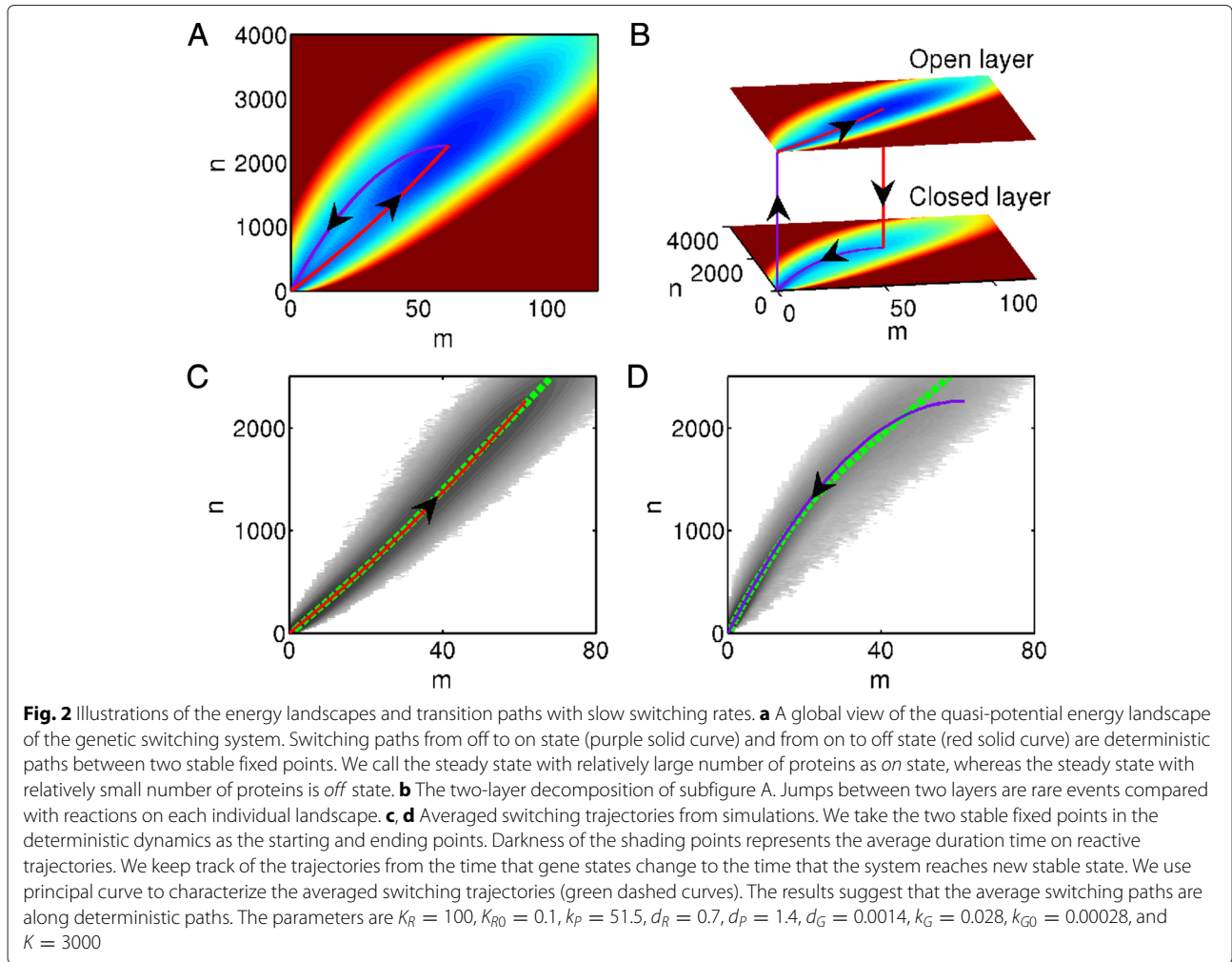
$$\begin{aligned} \dot{\mathbf{x}} &= \mathbf{c}(\alpha, \mathbf{x}(t)), \\ \dot{\boldsymbol{\sigma}} &= \boldsymbol{\sigma}(t) \mathbf{J}^T(\alpha, t) + \mathbf{J}(\alpha, t) \boldsymbol{\sigma}(t) + 2\mathbf{D}(\alpha, \mathbf{x}(t)), \end{aligned} \quad (4)$$

where $\mathbf{c}(\alpha, \mathbf{x})$ and $\mathbf{D}(\alpha, \mathbf{x})$ have the form

$$\begin{aligned} \mathbf{c} &= \begin{bmatrix} -d_R x + \alpha \left(k_{R0} + k_R \frac{y}{y+K} \right) \\ k_P x - d_P y \end{bmatrix}, \\ \mathbf{D} &= \frac{1}{2} \begin{bmatrix} d_R x + \alpha \left(k_{R0} + k_R \frac{y}{y+K} \right) & 0 \\ 0 & k_P x + d_P y \end{bmatrix}. \end{aligned} \quad (5)$$

$\mathbf{J}(\alpha, t)$ is the Jacobian of $\mathbf{c}(\alpha, \mathbf{x}(t))$, and $\boldsymbol{\sigma}(t)$ is the covariance matrix of the system. Thus, given $\alpha = 0$ or 1, the dynamics of the system can be well approximated by the time evolution of the moment Eqs. 4 and the transition paths mainly follow the solution of the ODEs (3) (Fig. 2a, b). This is verified by Gillespie's stochastic simulation algorithm and the results are shown in Fig. 2c, d. The darkness of the shading points represents the number of visits for transition paths, and the fitted curve shows nice coincidence with the deterministic path, which strongly supports our methodology.

This understanding can also be utilized to get a simple approximation of the invariant distribution. When the rates of genetic switches are extremely slow, the system will mainly stay in closed-state basin or open-state basin with rare transitions. So we have $\kappa \approx 0$, and we can approximate the steady state distribution P_α^{SS} by simple



Gaussian distribution $N(\mathbf{x}_\alpha^{ss}, \sigma_\alpha^{ss})$, where \mathbf{x}_α^{ss} and σ_α^{ss} are the corresponding steady state solution of Eq. 4. Furthermore we can combine the two distributions to depict the invariant distribution of the whole system through a Gaussian mixture model

$$P^{ss}(\mathbf{x}) \approx wN(\mathbf{x}_0^{ss}, \sigma_0^{ss}) + (1 - w)N(\mathbf{x}_1^{ss}, \sigma_1^{ss}), \quad (6)$$

where the mixture weight w is the proportion of time that the system stays on the closed layer. Since the mean time that gene stays on the open layer is d_G^{-1} , and the mean time that gene stays on the closed layer is approximately k_{G0}^{-1} (the steady state $O_p = 0$ on this layer), the value of w satisfies

$$\frac{w}{1 - w} \approx \frac{k_{G0}}{d_G}. \quad (7)$$

Solve it and we get $w = k_{G0}/(k_{G0} + d_G)$. This simple Gaussian mixture approximation is shown and compared with the accurate distribution in Fig. 3a, b. The approximation seems to fit the solutions of Eqs. 1-(2) well,

especially around the steady point. But some details corresponding with the switching paths are neglected, which makes some difference with the accurate solutions.

When the rates of genetic switches are relatively slow but not extremely slow, we can refine the approximation (Eq. 6) by taking into account the transition path information. On the open layer, we solve Eq. 4 with initial values $(\mathbf{x}_0^{ss}, \sigma_0^{ss})$ to get solutions $(\mathbf{x}_1^t, \sigma_1^t)$. Then we use the time-average of Gaussian distributions to approximate $P_1^{ss}(\mathbf{x})$, i.e.

$$P_1^{ss}(\mathbf{x}) \approx \frac{1}{T_1} \int_0^{T_1} p(\mathbf{x}, t) dt, \quad (8)$$

where $p(\mathbf{x}, t)$ is the probability distribution density of $N(\mathbf{x}_1^t, \sigma_1^t)$, and T_1 is the average time that system keeps staying on the open layer. The same approach applies to the approximation of $P_0^{ss}(\mathbf{x})$. Then we combine $P_0^{ss}(\mathbf{x})$ and $P_1^{ss}(\mathbf{x})$ to get

$$P^{ss}(\mathbf{x}) \approx wP_0^{ss}(\mathbf{x}) + (1 - w)P_1^{ss}(\mathbf{x}), \quad (9)$$

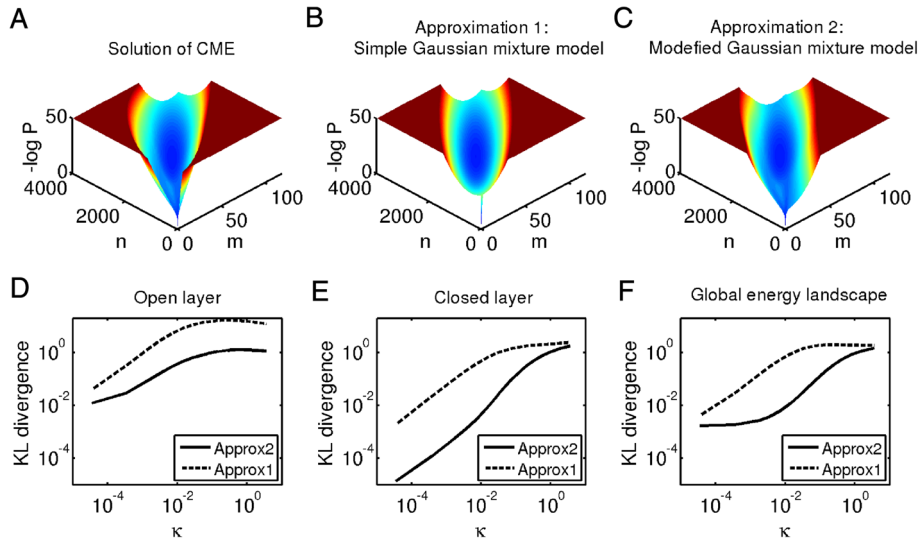


Fig. 3 Illustrations of the simple Gaussian mixture approximation and the modified Gaussian mixture approximation to the accurate energy landscape in the regime of slow switching rates. **a** The solution of CME which we take as the accurate energy landscape. **b** Illustration of the simple Gaussian mixture approximation. **c** Illustration of the modified Gaussian mixture approximation. **d, e, f** The relative errors of two approximation methods on the open layer, closed layer and global energy landscape. Dashed lines correspond to the simple Gaussian mixture approximation and solid lines correspond to the modified approximation. We use the KL divergence (i.e. Kullback-Leibler divergence) to measure the difference between two approximations and the accurate solution of CME. To make the KL divergence well defined, we assume that the minimum value of the distributions in the truncated domain is 10^{-16} (i.e. replace the values which are less than 10^{-16} with 10^{-16}). From the quantitative comparisons, we find that the modified approximation is superior than the simple approximation since more detailed information are encoded. Parameter values are the same with those in Fig. 2

where the value of w remains the same. We call Eq. 9 the modified Gaussian mixture approximation and compare it with the accurate distribution in Fig. 3c. Indeed, Eq. 6 is the limit of Eq. 9 when $\kappa \rightarrow 0$.

We also compute the relative errors of the two approximations and find that the modified Gaussian mixture approximation is superior than the simple Gaussian mixture approximation nearly in the whole slow regime (Fig. 3d, e, f). On both the open and the closed layers, the modified Gaussian approximation performs better than the simple Gaussian approximation. When the switching rates are extremely slow, the two approximations perform equally well. When κ is not very small but still belongs to the slow regime, the simple Gaussian approximation on the closed layer which is a single-point delta distribution, deviates far from the accurate solution. This is the main reason that simple Gaussian mixture approximation has relatively large errors.

Case B: fast regime

We now consider the case that genetic switches are much faster than the other reactions in the system. In this regime, the effective synthesis rate of mRNA is $(k_{RO}P + k_{R0})k_{on}/(k_{on} + k_{off})$ according to the QSSA, where k_{on} and k_{off} stand for the gene activation and inactivation rates.

The deterministic mean-field description of this model yields the ODEs:

$$\frac{dm}{dt} = \frac{(k_{R0} + k_R \frac{n}{n+K}) (k_{G0} + k_G \frac{n}{n+K})}{(d_G + k_{G0} + k_G \frac{n}{n+K})} - d_R m, \tag{10}$$

$$\frac{dn}{dt} = k_P m - d_P n.$$

This system has two stable fixed points and one saddle in physically reasonable parameter regime. These two stable fixed points correspond to the expressed and unexpressed states at which the copy number of proteins is at high or low state, respectively. As shown in the previous section, the global energy landscape of the slow regime also has two basins, which looks like the fast case in the first sight. But the essential nature is totally different. The bistability of the system in the slow regime originates from the simple two-layer dynamics induced by genetic switching, whereas the bistability in the fast regime is from an adiabatic reduction of the switching dynamics. The previous work of the authors has established a good framework to understand the metastability based on large deviation theory for Markov processes [15, 33]. In this framework, we assume that the steady state distribution $P^{SS}(\mathbf{x})$

satisfies $P^{ss}(\mathbf{x}) \propto \exp(VS(\mathbf{x}))$, where $S(\mathbf{x})$ is the global quasi-potential function that we focus on, and the volume size V is inversely proportional to the noise strength. From WKB asymptotics [13] or the large deviation theory [15, 33], we know that $S(\mathbf{x})$ satisfies the Hamilton-Jacobi equation $H(\mathbf{x}, \nabla S) = 0$ from the leading order analysis for the corresponding CME, where H is the Hamiltonian of the system. However, solving the Hamilton-Jacobi equation is not an easy task, and we utilize an alternative efficient optimization approach to obtain $S(\mathbf{x})$. We define the local quasi-potential $S(\mathbf{x}; \mathbf{x}_0)$ with respect to a metastable state \mathbf{x}_0 as

$$S(\mathbf{x}; \mathbf{x}_0) = \inf_{T>0} \inf_{\phi, \phi(0)=\mathbf{x}_0, \phi(T)=\mathbf{x}} \int_0^T L(\phi, \dot{\phi}) dt, \quad (11)$$

where L is the Lagrangian, which is the Legendre transform of the Hamiltonian H , and the path ϕ are all possible continuous connections from \mathbf{x}_0 to \mathbf{x} within time T . From classical mechanics, the double minimization problem (11) can be transformed to a single minimization with

respect to the intrinsic representation of the curves $\phi(s)$ by Maupertuis' principle [15, 34] as

$$S(\mathbf{x}; \mathbf{x}_0) = \inf_{\substack{\phi(0)=\mathbf{x}_0, \phi(1)=\mathbf{x} \\ H(\phi, \mathbf{p})=0, H_{\mathbf{p}}(\phi, \mathbf{p})=\lambda \phi'}} \int_0^1 p(s) \cdot d\phi(s), \quad (12)$$

where $p(s)$ and λ are determined from the constraints for each $\phi(s)$. This minimization problem can be efficiently solved by using the geometric minimum action method (gMAM) [15, 34]. $S(\mathbf{x})$ can be obtained from its local version $S(\mathbf{x}; \mathbf{x}_0)$ by a suitable sticking procedure [15, 33]. With this approach, we can construct the global quasi-potential energy landscape $S(\mathbf{x})$ and obtain the most probable transition paths between two metastable states. More details can be referred to [15].

With proper values of parameters, the system has two metastable states and the global quasi-potential energy landscape can be constructed as shown in Fig. 4a. The landscape looks similar with Fig. 2, but transition paths between two basins show different features. In the slow

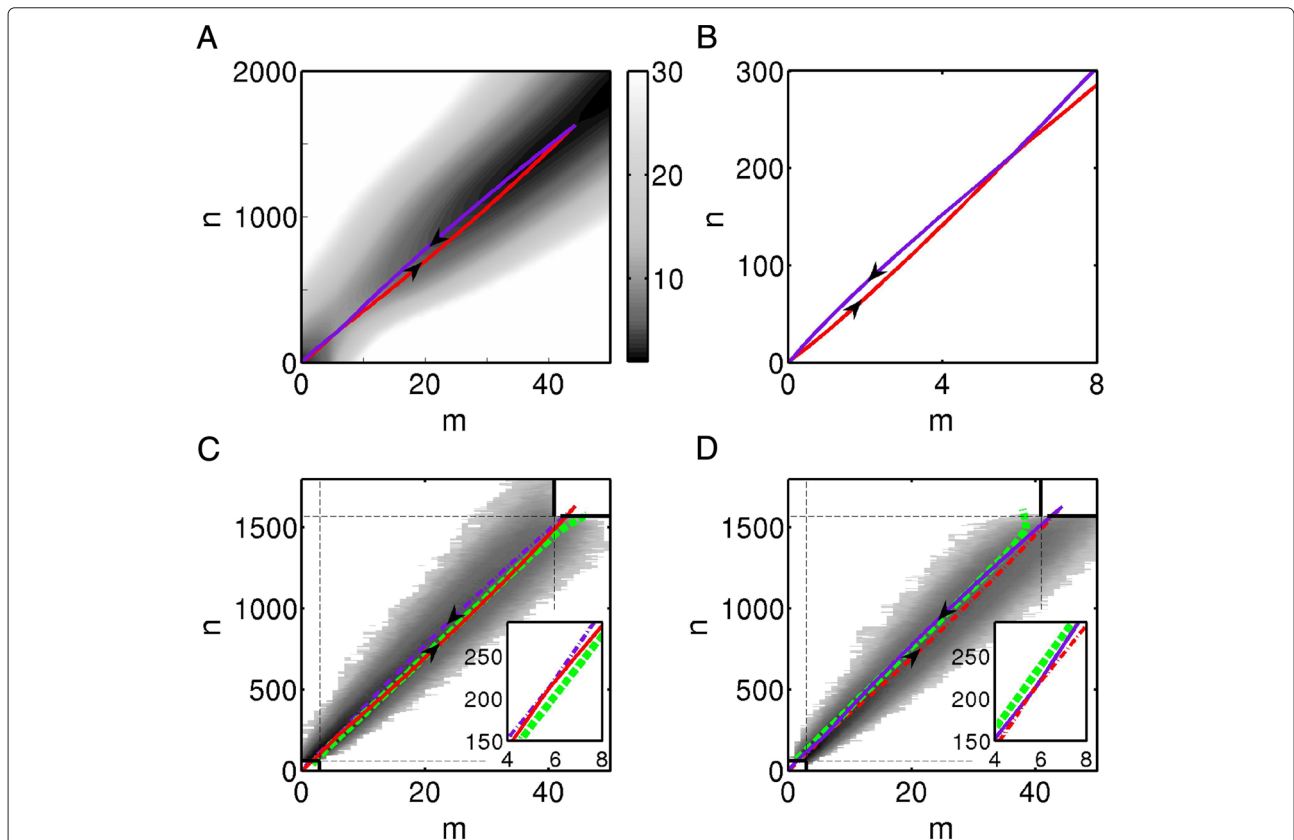


Fig. 4 Illustrations of the energy landscapes and transition paths with fast switching rates. **a** The global quasi-potential energy landscape of the genetic switching system. The most probable switching paths from off to on state (purple solid curve) and from on to off state (red solid curve) both pass through the saddle point. **b** Details of subfigure A. **c, d** The principal curves (green dashed curves) fitted from shading points obtained by Gillespie's algorithm compare with the optimal transition paths. The result shows good matching between the fitted curves and the theoretical results. The parameters $d_G = 140$, $k_G = 2800$, $k_{G0} = 28$, and the others are the same with those in Fig. 2

regime, the transition paths are characterized by the solutions of deterministic ODEs (3) and they do not intersect at any point. In the fast regime, the most probable transition paths of on-to-off switching and off-to-on switching are not identical, either. They intersect at a unique point, i.e. the saddle point of the deterministic ODEs (10), but they do not cross over each other (Fig. 4a, b). We also present the results by Gillespie's algorithm and compare the theoretically predicted optimal transition paths and the fitted curves in Fig. 4c, d. The comparison shows good matching between the fitted curves and the theoretical transition paths. Our results are consistent with the previous studies [15, 35].

Case C: intermediate regime and comparisons

In both the fast and slow regimes, we have already illustrated some analytical approaches to quantitatively understand the global energy landscape and transitions between metastable states in gene expression model with positive feedback. But in the intermediate regime, where all of the reactions are on the same time scale, these analytical methods fail, since it is difficult to find the appropriate deterministic equations and hard to know whether steady states exist in the system. Thus we compute the key properties in the intermediate regime through Gillespie's algorithm and compare the results with the slow and fast regimes.

We simulate the mean switching time (MST) with different parameters. The MST is the average transition time between two basins on the global energy landscape, which is often used to characterize the stability of the attractors. Figure 5 shows the MST with respect to different κ and positive feedback strength K . When κ increases from 10^{-5} to 10^5 , the gene activation rate changes from slow to fast regimes. Since we focus on switching times

between metastable states, the value of K should be chosen to guarantee that bistability always exists in the system with respect to all the possible values of κ . We define the existence of bistability as that the system stay at each metastable state with probability more than 0.01. In the slow regime, bistability always exists for any positive K . In the intermediate and fast regimes, we find two boundary values of K , which are $K = 2754$ and $K = 3211$. When $K < 2754$, the system will stay at the off state with probability less than 0.01, whereas when $K > 3211$, the system will stay at the on state with probability less than 0.01. The smaller K is, the larger positive feedback strength is.

In the slow regime, the MST can be characterized by the waiting time for transitions to occur, so the MST decreases as κ increases. In this regime, K has little effect on MST, which is consistent with our analysis in Case A that the MST of on-to-off and off-to-on transitions is approximately d_G^{-1} and k_{G0}^{-1} , respectively. When κ increases to a value between 10^{-2} and 10^{-1} , the MST decreases to its global minimum point, and a neighborhood of this point is corresponding to the intermediate regime. Then the MST goes up with κ increasing, and finally reaches a stable value in the fast regime. In the fast regime, when K changes from 2754 to 3211 (i.e. the strength of positive feedback decreases), the MST of off-to-on switch increases by a factor of about 10, whereas the MST of on-to-off switch decreases by a factor of about 10^3 . Thus, if the positive feedback becomes stronger, the system will be more probable to stay at the on state, with shorter off-to-on switching time and longer on-to-off switching time.

Longer MST indicates that metastable states are more stable. Thus, our results in Fig. 5 show that the slow and fast regimes perform better in stability, whereas in the intermediate regime, genetic switches, transcription and

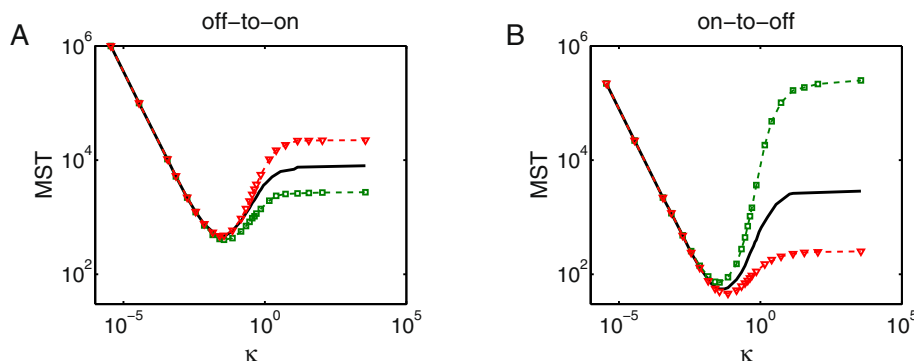


Fig. 5 The mean switching time (MST) as a function of κ with different positive feedback strength. **a** MST curve of off-to-on switch. **b** MST curve of on-to-off switch. Parameters are the same with with Fig. 2 except d_G , k_G , k_{G0} , and K . When κ changes, d_G should change simultaneously to keep d_G/k_G , d_G/k_{G0} locked (these ratios are the same with Fig. 2). We use different strength of positive feedback for simulations, which is $K = 2754$ (green dashed lines with square markers), 3000 (black solid lines), 3211 (red dashed lines with triangular markers). All of the results are obtained from Gillespie's algorithm

translation are on the same time scale, which contributes to large fluctuations of molecule numbers and relatively short MST. These properties remain the same with different positive feedback strength. Furthermore, we also did some interesting simulations to study the change of transition paths as a function of κ (see Additional file 1 for details).

We should mention that the above points are direct observations from numerical results, which are hard to be proven analytically. However, we can confirm from the derivations in Case A that when κ is extremely small, the MST of on-to-off and off-to-on transitions is approximately d_G^{-1} and k_{G0}^{-1} . That is to say, the downward-sloping shapes of the left parts of the MST curves are qualitatively robust. Besides, when κ goes to infinity, the MST will converge to some constant values [15]. So the flat parts on the right of the MST curves are also theoretically supported. The existence of “valleys” in the MST curves is in accordance with the intuition, but it is difficult to give a rigorous proof. To investigate the robustness of the whole qualitative shape of the MST curve with respect to different feedback actions, we simulated all the results with $O_P = n^2/(n^2 + K^2)$, and found that the shapes of the MST curves are almost the same as in Fig. 5. We may reasonably infer that all results still hold assuming that O_P is a Hill function and the bistability exists in both slow and fast regimes (see Additional file 1 for more details). These results suggest a natural methodology to distinguish the different genetic activation/inactivation rates in the cellular phenotype observations, which is described in detail in the continued text.

Distinguishing the three regimes

In molecular experiments, it is hard to directly measure the switching rates of the active or inactive gene state. Based on the obtained results, we propose a method to distinguish which regime the considered system falls in. Consider a gene expression circuits with the positive feedback in a eukaryotic cellular system (Fig. 1), such as the yeast or the immune cell etc, we can observe the level of proteins through fluorescence or other markers by flow cytometry. Below we will show how to take advantage of these perceivable data and related analysis to make corresponding distinction.

First let us introduce how to identify the intermediate regime. We utilize the strength of cellular memory as the index, and more concretely, we sort the cells by the protein level and observe how long the sorted cells will take to reach the original invariant distribution of proteins. The detailed flow cytometry procedure of our experiments *in silico* is stated in the following two steps.

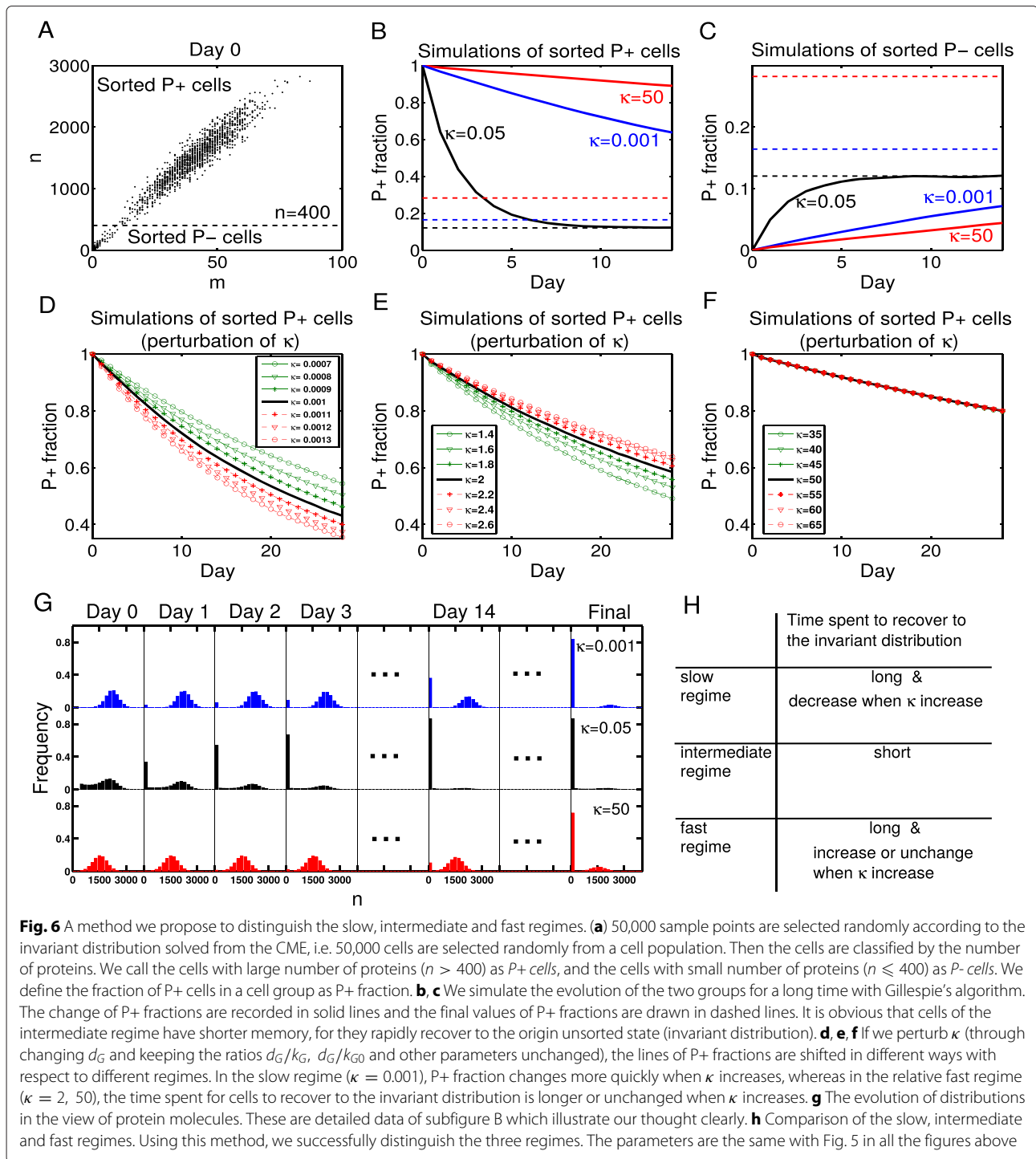
- Step 1: Sorting the cells by the protein levels. For example, we choose $n = 400$ as the threshold of

classification in our model setup, where n is the number of protein molecules in a cell. We call the cells with large number of proteins ($n > 400$) as $P+$ cells, and the cells with small number of proteins ($n \leq 400$) as $P-$ cells. We define the fraction of $P+$ cells in a cell group as the $P+$ fraction. Figure 6a illustrates the numerically sorted results of the system, including 50,000 sample points which are randomly selected according to the invariant distribution, and $P+/P-$ cells are classified by the protein level in each cell. To apply this approach to the gene regulatory network with different parameters, one need to choose an appropriate threshold according to the real situation, which needs to lie between two metastable states.

- Step 2: Culturing the sorted $P+$ and $P-$ groups of cells separately, and recoding the changes of $P+$ fractions in each group. Figure 6b, c show the simulation results of the $P+$ and $P-$ groups respectively. The evolution of the protein distributions of the sorted $P+$ group is simulated and illustrated in Fig. 6g, including results of the slow, intermediate and fast regimes. The sorted $P+$ group of cells in the intermediate regime can be observed to evolve quickly to its initial equilibrium state. In the third day, a large percent of $P+$ cells in the intermediate regime have evolved to the $P-$ group, whereas the $P+$ cells in the slow and fast regimes have little change. This suggests that cellular memory is much stronger in the slow and fast regimes compared with the intermediate regime.

In the real experiments, we do not know whether the time for an appreciable change of protein levels is long or short without reliable reference. But the MST in the intermediate regime is approximately the time for cells to reach a metastable state after leaving another metastable state, thus we can use it as a reference time. We define T as the on-to-off (or off-to-on) transition time, which does not contain the dwell time at the on (or off) state. Since the synthesis and degradation rates of mRNAs and proteins can be estimated through experimental approaches, we can theoretically figure out the time that the system will spend to reach the off state after leaving the on state. That is to say, T can be roughly estimated a priori. So we make experiments like Fig. 6g for a period of time which is close to T . If the value of $P+$ fraction changes evidently towards the equilibrium level, we will know that the system belongs to the intermediate regime.

Second we try to identify the fast and slow regimes through the perturbation-response behavior of changing κ . As shown in Fig. 5a, the MST becomes larger and larger when κ approaches to zero, but it increases and gradually stays close to a fixed value when κ is large. Thus, if we perturb κ through injection of particular



chemicals, cells in the two regimes will perform differently in the experiments we designed above. Figure 6d, e, f illustrate the results of corresponding numerical simulations. In the slow regime ($\kappa = 0.001$), *P+* fraction changes more quickly when κ increases, whereas in the relatively fast regime ($\kappa = 2, 50$), the time for cells to recover to the invariant distribution is longer

or unchanged when κ increases. With this perturbation-response approach, the slow and fast regimes can also be distinguished from each other (Fig. 6h). When the positive feedback strength K takes the boundary values mentioned in case C, the simulation experiments above still work well (see Additional file 1 for details). This proposal, based on our theoretical understanding of the

feature of the three regimes, is yet to be justified by experimentalists.

Discussion

Understanding the dynamical features and the energy landscapes of the gene regulatory network is an interesting and challenging topic in quantitative biology. Some approaches focused on the regime of fast switching rates have been developed in recent years, whereas the slow regime which many eukaryotic cells correspond to attracts less attention. In this paper, besides the theoretical analysis of the essential dynamics in the slow, intermediate and fast regimes, we provide a method to distinguish these three regimes. The result is meaningful and enlightening for the design of corresponding biological experiments.

With regard to a previous work about stochastic dynamics of single regulating genes [19], we provide a systematic study to understand the genetic switching dynamics when the switching rates are in different regimes. Although the global quasi-potential energy landscape and other properties of the system can be obtained from brutal force through numerical simulations, the theoretical reductions can reveal the intrinsic mechanisms. We analyze the model theoretically in the slow and fast regimes and successfully reconstruct the global energy landscape which reflects the inherent pattern of system behavior. But we obtain little theoretical results in the intermediate regime, since the nonexistence of time-scale separation makes this problem more challenging. Recent new ideas proposed in this field may be helpful in the study of this regime [36, 37].

A related but different concept with multi-layer landscapes for the DNA fast switching case was mentioned in a recent literature [36]. However, we are mainly concerned with the two-layer case when the system is in slow regime. Besides understanding the system as decoupled layers, we further study the properties about transition paths and average switching time, and we construct the global potential energy landscape through simple Gaussian mixture and modified Gaussian mixture approximations. These pictures are quite different from the case of fast regime.

The MST is a key characterization of gene regulatory networks and has been studied analytically with respect to different genetic switching rates in [38]. Compared with the detailed theoretical analysis for a typical one-dimensional self-activating switch model, we can not give fully theoretical results in the fast regime, i.e. the strictly adiabatic regimes considered in [38], since our model involves both mRNA and protein explicitly. However, we present efficient numerical methods to make the quantitative estimates. The results for the extremely slow regime are also consistent with those obtained for non-adiabatic

regime (i.e. Eq. 7). We also suggested the landscape as an intuitive tool to understand the essential dynamics of our model. Overall, the shape of the MST curve we obtained as a function of κ is qualitatively similar as in [38] as well.

The proposed quasi-potential energy function $S(\mathbf{x})$ in the fast regime has close connection with the potential function considered in previous studies. In Wang et al's approach [14, 39], they proposed to define the energy landscape as $U(\mathbf{x}) = -\ln P^{ss}(\mathbf{x})$, where $P^{ss}(\mathbf{x})$ is the steady state distribution. We have the connection that $S(\mathbf{x}) = \lim_{V \rightarrow \infty} V^{-1}U(\mathbf{x})$ [15, 40] since $U(\mathbf{x})$ itself depends on the system size V . We also note that the proposed quasi-potential $S(\mathbf{x})$ satisfies the same Hamilton-Jacobi equation as the potential defined in [41, 42] in the diffusive approximations. But in addition, we provided a variational minimization approach to explicitly construct the $S(\mathbf{x})$. Nonetheless, this construction is only valid for the fast regime and it is quite different from the picture we showed in the slow regime.

Although our four-state genetic switching model is based on chromatin remodeling, this is no longer expressed in the final simplified two-state model. In fact, some prokaryotic cells also have similar gene regulatory networks like Fig. 1b [17]. Our study mainly focuses on the two-state model, which can be suitable for a wide variety of cellular systems rather than the motivating chromatin remodeling process. Our simulation experiment to distinguish the slow, intermediate and fast regimes is an important point of the paper. So far, we are not able to perform real experiments to make the verification. The real experimental systems may be also more complicated than the model considered here. But our simulation experiment is logically true and self-consistent, and there have been related experiments performing well in Th2 cells [16]. Finding suitable system and validating our proposal will be an interesting task in the future.

Our study has potential application in single-cell experiments [43–45]. In the single-cell gene expression analysis, the protein level of each cell at different time points can be obtained through tracking and recording the behavior of each cell by fluorescence microscopy. With these data, one can obtain the switching time as the time difference between two consecutive states of a cell. Thus the MST can be roughly estimated. Similar as the procedure in flow cytometry experiments, the information about MST can be utilized to identify the intermediate regime, and the fast and slow regimes can be distinguished by perturbing κ . However, these points remain to be investigated in detail in the follow-up researches.

Conclusions

In this paper, we proposed a methodology to distinguish the regime of genetic switching rates from phenotypic variations in a simulation experiment. To accomplish this

task, we put forward different reduction and investigation methods to reveal the essential dynamics of the genetic switching system when the switching rates are relatively slow, fast, or medium compared with the degradation rates of proteins. Although in the fast and slow regimes, the observed steady state distributions of the mRNAs and proteins are similar, we showed that the underlying mechanisms are qualitatively different. Based on our results, we provided a simulation experiment to distinguish these three regimes. The intermediate regime can be identified from the fact that the strength of cellular memory is lower than the other two cases, and the fast and slow regimes can be distinguished by their different perturbation-response behavior with respect to the switching rates perturbations. We also discussed the robustness of our results with respect to different choices of parameters and feedback function forms. This study provides insights to the experimental analysis of the gene expression system. It will be interesting to investigate similar problems in different experimental setups in the future.

Additional file

Additional file 1: Supporting information. Simplification of the four-state genetic switching model; Gaussian approximation in the regime of slow switching rates; An extension of the results: $O_p = n^k / (n^k + K^k)$, $k \in \mathbb{N}$; Distinguishing the three regimes when the positive feedback strength K takes the boundary values; The change of switching paths as a function of κ ; and biological relevance of the parameter values.

Abbreviations

MST: Mean switching time; QSSA: Quasi-steady-state approximation; CME: Chemical master equation; gMAM: Geometric minimum action method.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

TL and FL designed the research. YC and CL performed the research. All authors analyzed the data and wrote the paper. All authors read and approved the final manuscript.

Acknowledgements

The work is supported by NSFC grants no. 11174011, 11021463 (F.Li), 11171009, 11421101 and 91130005 and the National Science Foundation for Excellent Young Scholars (Grant No. 11222114) (T.Li). The authors thank Professor Jin Wang for helpful discussions.

Author details

¹LMAM and School of Mathematical Sciences, Peking University, No. 5 Yiheyuan Road, 100871 Beijing, China. ²School of Physics, Peking University, No. 5 Yiheyuan Road, 100871 Beijing, China. ³Center for Quantitative Biology, Peking University, No. 5 Yiheyuan Road, 100871 Beijing, China.

Received: 22 December 2014 Accepted: 29 May 2015

Published online: 18 June 2015

References

- Swain PS, Elowitz MB, Siggia ED. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc Natl Acad Sci USA*. 2002;99:795–12800.
- Eldar A, Elowitz MB. Functional roles for noise in genetic circuits. *Nature*. 2010;467:167–73.
- Balázsi G, van Oudenaarden A, Collins JJ. Cellular decision making and biological noise: From microbes to mammals. *Cell*. 2011;144:910–25.
- Kussell E, Leibler S. Phenotypic diversity, population growth, and information in fluctuating environments. *Science*. 2005;309:2075–8.
- Acar M, Mettetal JT, van Oudenaarden A. Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet*. 2008;40:471–5.
- Choi PJ, Cai L, Frieda K, Xie XS. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*. 2008;322:442–6.
- Norman TM, Lord ND, Paulsson J, Losick R. Memory and modularity in cell-fate decision making. *Nature*. 2013;503:481–6.
- Maamar H, Raj A, Dubnau D. Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science*. 2007;317:526–9.
- Gandhi SJ, Zenklusen D, Lionnet T, Singer RH. Transcription of functionally related constitutive genes is not coordinated. *Nat Struct Mol Biol*. 2010;18:27–34.
- Stewart-Ornstein J, Weissmanemail JS, El-Samademail H. Cellular noise regulons underlie fluctuations in *Saccharomyces cerevisiae*. *Mol Cell*. 2012;45:483–93.
- Raser JM, O'Shea EK. Control of stochasticity in eukaryotic gene expression. *Science*. 2004;304:1811–4.
- Munsky B, Neuert G, van Oudenaarden A. Using gene expression noise to understand gene regulation. *Proc Natl Acad Sci USA*. 2012;336:183–7.
- Assaf M, Roberts E, Luthey-Schulten Z. Determining the stability of genetic switches: explicitly accounting for mRNA noise. *Phys Rev Lett*. 2011;106:248102.
- Wang J, Zhang K, Wang E. Kinetic paths, time scale, and underlying landscapes: A path integral framework to study global natures of nonequilibrium systems and networks. *J Chem Phys*. 2010;133:125103.
- Lv C, Li X, Li F, Li T. Constructing the energy landscape for genetic switching system driven by intrinsic noise. *PLoS ONE*. 2014;9:88167.
- Mariani L, Schulz EG, Lexberg MH, Helmstetter C, Radbruch A, Löhning M, Höfer T. Short-term memory in gene induction reveals the regulatory principle behind stochastic *il-4* expression. *Mol Syst Biol*. 2010;6:359.
- Hao N, Budnik BA, Gunawardena J, O'Shea EK. Tunable signal processing through modular control of transcription factor translocation. *Science*. 2013;339:460–4.
- de la Serna IL, Ohkawa Y, Imbalzano AN. Noise contributions in an inducible genetic switch: A whole-cell simulation study. *Nat Rev Genet*. 2006;7:461–73.
- Feng H, Han B, Wang J. Adiabatic and non-adiabatic non-equilibrium stochastic dynamics of single regulating genes. *J Phys Chem B*. 2011;115:1254–61.
- Ackers GK, Johnson AD, Shea MA. Extinction of metastable stochastic populations. *Proc Natl Acad Sci USA*. 1982;79:1129–33.
- Paulsson J, Berg OG, Ehrenberg M. Stochastic focusing: Fluctuation-enhanced sensitivity of intracellular regulation. *Proc Natl Acad Sci USA*. 2000;97:7148–153.
- Bodenstein M. Eine theorie der photochemischen reaktionsgeschwindigkeit. *Z Phys Chem*. 1913;85:329–97.
- Chapman DL, Underhill LK. The interaction of chlorine and hydrogen. the influence of mass. *J Chem Soc Trans*. 1913;103:496–508.
- Metzler R, Wolynes PG. Number fluctuations and the threshold model of kinetic switches. *Chem Phys*. 2002;284:469–79.
- Sasai M, Wolynes PG. Stochastic gene expression as a many-body problem. *Proc Natl Acad Sci USA*. 2003;100:2374–379.
- Paulsson J. Summing up the noise in gene networks. *Nature*. 2004;427:415–8.
- Thattai M, van Oudenaarden A. Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA*. 2001;98:8614–619.
- Waddington CH, Kacser H. *The Strategy of the Genes: A Discussion of Some Aspects of Theoretical Biology*, 1st edn. London: George Allen and Unwin; 1957.
- Qian H, Reluga TC. Nonequilibrium thermodynamics and nonlinear kinetics in a cellular signaling switch. *Phys Rev Lett*. 2005;94:028101.
- Kampen NGV. A power series expansion of the master equation. *Can J Phys*. 1961;39:551.
- Kampen NGV. The expansion of the master equation. *Adv Chem Phys*. 1976;34:245.

32. Kubo R, Matsuo K, Kitahara K. Fluctuation and relaxation of macrovariables. *J Stat Phys*. 1973;9:51–96.
33. Lv C, Li X, Li F, Li T. Energy landscape reveals that the budding yeast cell cycle is a robust and adaptive multi-stage process. *PLoS Comput Biol*. 2015;11:1004156.
34. Heymann M, Vanden-Eijnden E. The geometric minimum action method: A least action principle on the space of curves. *Comm Pure Appl Math*. 2008;61:1052–117.
35. Assaf M, Meerson B. Extinction of metastable stochastic populations. *Phys Rev E*. 2010;81:021116.
36. Zhang K, Sasai M, Wang J. Eddy current and coupled landscapes for nonadiabatic and nonequilibrium complex system dynamics. *Proc Natl Acad Sci USA*. 2013;110:14930–35.
37. Zhang B, Wolynes PG. Stem cell differentiation as a many-body problem. *Proc Natl Acad Sci USA*. 2014;111:10185–90.
38. Walczak AM, Onuchic JN, Wolynes PG. Absolute rate theories of epigenetic stability. *Proc Natl Acad Sci USA*. 2005;102:18926–31.
39. Wang J, Xu L, Wang E. Potential landscapes and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. *Proc Natl Acad Sci USA*. 2008;105:12271–76.
40. Freidlin MJ, Wentzell AD. *Random Perturbations of Dynamical Systems*, 2nd edn. New York: Springer; 1998.
41. Ao P. Potential in stochastic differential equations: novel construction. *J Phys A: Math Gen*. 2004;37:25–30.
42. Xu S, Jiao S, Jiang P, Ao P. Two-time-scale population evolution on a singular landscape. *Phys Rev E*. 2014;89:012724.
43. Fritzsche FSO, Dusny C, Frick O, Schmid A. Single-cell analysis in biotechnology, systems biology, and biocatalysis. *Annu Rev Chem Biomol Eng*. 2012;3:129–55.
44. Wang D, Bodovitz S. Single cell analysis: the new frontier in 'omics'. *Trends Biotechnol*. 2010;28:281–90.
45. Han Q, Bagheri N, Bradshaw EM, Hafler DA, Lauffenburger DA, Love JC. Polyfunctional responses by human t cells result from sequential release of cytokines. *Proc Natl Acad Sci USA*. 2012;109:1607–12.

**Submit your next manuscript to BioMed Central
and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

