Biophysical Journal, Volume 98

Supporting Material

The potential landscape of genetic circuits imposes the arrow of time in stem cell differentiation

Jin Wang, Li Xu, Erkang Wang, and Sui Huang

Supporting Information for The "potential" landscape of genetic circuits imposes the arrow of time in stem cell differentiation

Jin Wang [ab*] Li Xu [a] Erkang Wang [a*] ,Sui Huang [c*]

[a] State Key Laboratory of Electroanalytical Chemistry
Changchun Institute of Applied Chemistry
Chinese Academy of Sciences
Changchun, Jilin 130022 People's Republic of China
[b] Department of Chemistry & Department of Physics
State University of New York at Stony Brook
Stony Brook, NY 11794
[c]Department of Biological Sciences and Institute for
Biocomplexity and Informatics, Biological Sciences Building, 2500
University Dr. N.W. Calgary, Alberta, Canada T2N 1N4;
Children's Hospital, Harvard Medical School,
1 Blackfan Circle, Boston, MA 02115, USA

[*] Corresponding Author: E-mail: jin.wang.1@stonybrook.edu, ekwang@ciac.jl.cn, sui.huang@ucalgary.ca

1 Potentials in non-equilibrium, non-integrable systems

In the gene circuit (Eq. (1)and(2)):

$$d\mathbf{x}/dt = \mathbf{F}(\mathbf{x}) \tag{1}$$

the vector $\mathbf{F}(\mathbf{x})$ is the force that drives the system. However, the force $\mathbf{F}(\mathbf{x})$ cannot in general be written as a gradient of a potential U, $\mathbf{F}(\mathbf{x}) \neq -grad(U)$ for systems of more than one dimensions. In other words, $\mathbf{F}(\mathbf{x})$ is not a pure gradient of a potential, but there is another force, $\mathbf{F}_c$ stemming which contributes to the dynamics:

$$\mathbf{F}(\mathbf{x}) = \mathbf{F}_c - \mathbf{D} * grad(U) \tag{2}$$

where $\mathbf{D}$ is the diffusion coefficient tensor. It is not easy to separate these two components of the driving force, i.e., to calculate $\mathbf{F}_c$ and $grad(U)$. However, in a stochastic system in which each state $\mathbf{x}$ is described probabilistically, the information about the probability in time and in steady state to find the system in state space position $\mathbf{x}$, $P(\mathbf{x})$ can be found by either solving the corresponding probabilistic equation (master equation or diffusion equation) or through Monte Carlo simulations. We would like to explore how information on the steady state probability distribution will correlate with the dynamics of the system.

In the equilibrium case, the potential and equilibrium probability is linked by the Boltzman law:

$$U(\mathbf{x}) \sim -ln(P_{ss}(\mathbf{x})) \tag{3}$$

The question is how the two quantities, $\mathbf{F}_c$ and U are connected in non-equilibrium systems like gene regulatory circuits. The statistical description is justified because of the stochastic fluctuations in $\mathbf{x}$, or "gene expression

noise" [1, 2]. The effect of such molecular noise on gene circuits has been widely studied theoretically and experimentally where the probability distribution $P(\mathbf{x})$ is often estimated from snapshot measurements of xi in random ensembles of gene circuits, i.e., in cell populations. We can now, instead of the deterministic dynamical equations, explore the corresponding probabilistic dynamics which is governed by the Fokker-Planck equation for the temporal evolution of the probability distribution $P(\mathbf{x})$:

$$\begin{aligned}\frac{\partial P(x_1,x_2,t)}{\partial t} &= -\frac{\partial}{\partial x_1}[F_1(x_1,x_2)*P] - \frac{\partial}{\partial x_2}[F_2(x_1,x_2)*P] \\ &+\frac{\partial^2}{\partial {x_1}^2}[D(x_1,x_2)*P] + \frac{\partial^2}{\partial {x_2}^2}[D(x_1,x_2)*P] \qquad (4)\end{aligned}$$

Eq.(4) essentially describes the probability P for finding the circuit state $S = \mathbf{x}(t)$ in the state space as it moves driven by the drift (first two terms on the right hand side in Eq. (4)), representing the interactions defined by F in Eq (1), and by noise that can be represented as "diffusion in state space" (last two terms on the right hand side in (4)). For simplicity, D is chosen to be a constant.

It is more intuitive to describe the evolution of the probability density in terms of the probability flux $\mathbf{J}(\mathbf{x},t)$ in $\mathbf{x}$ at t. Thus, because of continuity, $\frac{\partial P}{\partial t} = -\nabla \cdot J(\mathbf{x},t)$, where $J(\mathbf{x},t) = \mathbf{F} * P - \mathbf{D} * \frac{\partial}{\partial \mathbf{x}} P$. At steady state $\frac{\partial P_{ss}}{\partial t} = -\nabla \cdot J_{ss}(\mathbf{x},t) = 0$. While the divergence of the steady state probability flux J_{ss} , $J_{ss}(\mathbf{x},t)$, must vanish (divergence free), the flux J_{ss} itself does not need to vanish in non-equilibrium systems. This is because due to the constraints in the non-equilibrium system, detailed balance does not necessarily hold. Only for equilibrium systems, where the detailed balance is satisfied, is $J_{ss} = 0$. Thus, from the definition of $\mathbf{J}(\mathbf{x},t)$ one can write for the steady state

$$\mathbf{F} = \mathbf{J}/P_{ss} + \mathbf{D} * grad(P_{ss})/P_{ss}$$

$$
\begin{aligned}
&= \mathbf{J}/P_{ss} - \mathbf{D} * grad[-ln(P_{ss})] \\
&= \mathbf{F}_c - \mathbf{D} * grad(U) \qquad (5)
\end{aligned}
$$

Here, $\mathbf{F}_c = \mathbf{J}/P_{ss}$ reflects the additional force linking to the non-vanishing flux, and Fc stands for curl force. The probability flux is divergence free due to the steady state conditions. Divergence free flux has no sinks or leaks, and therefore, does not have a place to start or end. It is in this sense that the flux has a curl nature, rotating around. Importantly, in Eq.(5), we have decomposed the force driving the dynamics of the system into two terms [3]. One is the curl force $\mathbf{F}_c$ and the other is the gradient of the potential U where U is linked with the steady state probability $grad(U) = -grad[ln(P_{ss})]$. This allows us to naturally introduce the non-equilibrium landscape U as $U \sim -ln(P_{ss})$ similar to the equilibrium situation, Eq.(3). The difference between equilibrium systems (i.e., protein folding) and general non-equilibrium systems (i.e., regulatory circuits or networks) is that although the potential is linked to the steady state probability in a similar way in both cases, the dynamics of the equilibrium system follows a gradient of the potential whereas the dynamics of the non-equilibrium systems is governed by both the gradient of the potential as well as the curl flux [3]. Thus, the major difference between equilibrium and non-equilibrium dynamics lies in how the potential (or steady state probability) is linked to the dynamics.

2 Solving the diffusion equation

We use basic finite difference methods for approximating the solutions to the diffusion equations. A two-dimensional square space is divided into square lattice boxes with a space h and the mesh points $(x_{1,i}, x_{2,j}) = (ih, jh), i(j) = 0, 1, ..., N.$ within the region. The diffusion equation makes use of first and second differences in the t direction. Let τ denotes the length of a time step, so that $t = t_k = k\tau(k = 0, 1, ..., M)$. Then $P_{i,j}^k$ denotes the P value on mesh

point $(x_{1,i}, x_{2,j})$ at time t_k , and $F^k_{1,i,j}$ denotes the F_1 value on mesh point $(x_{1,i}, x_{2,j})$. We used a difference scheme that corresponds to Euler's method for the equation:

$$\begin{aligned}
\frac{P^{k+1}_{i,j} - P^k_{i,j}}{\tau} &= -P^k_{i,j}\left(\frac{F^k_{1,i+1,j} - F^k_{1,i-1,j}}{2h} + \frac{F^k_{2,i,j+1} - F^k_{2,i,j-1}}{2h}\right) \\
&\quad -F^k_{1,i,j}\frac{P^k_{i+1,j} - P^k_{i-1,j}}{2h} - F^k_{2,i,j}\frac{P^k_{i,j+1} - P^k_{i,j-1}}{2h} \\
&\quad +D\left(\frac{P^k_{i+1,j} - 2P^k_{i,j} + P^k_{i-1,j}}{h^2}\right. \\
&\quad \left.+\frac{P^k_{i,j+1} - 2P^k_{i,j} + P^k_{i,j-1}}{h^2}\right) \\
&= Y(x_{1,i}, x_{2,j}, t_k) \qquad (6)
\end{aligned}$$

Starting with the initial conditions which we can choose, we can step from any value of t to $t + \tau$ with $P(x_1, x_2, t + \tau) = P(x_1, x_2, t) + \tau Y(x_1, x_2, t)$ for all of the mesh points x_1, x_2 in the region. The boundary conditions provide the values on the boundary or outside the region. We obtain almost identical results by using both the Neumann and Dirichlet boundary conditions. This method is explicit because each new value of P can be computed directly from values of P at the previous time step. More complicated methods are implicit because they involve the solution of systems of equations at each step. If we now bring all the P^{k+1} to the left-hand side and use the vector notation

$$P^k = \begin{bmatrix} P^k_{11} & P^k_{12} & .. & P^k_{1N} \\ P^k_{21} & P^k_{22} & .. & P^k_{2N} \\ : & : & .. & : \\ P^k_{N1} & P^k_{N2} & .. & P^k_{NN} \end{bmatrix} \qquad (7)$$

we obtain the value of P by solving the equation of the form

$$AP^{k+1} = CP^k \qquad (8)$$

We solved the diffusion equations with the help of Vcell software [4]. In the present case we used the square region with $0 \leq x_1 \leq 3$ and $0 \leq x_2 \leq 3$, and the Neumann boundary condition. If the time step exceeded a critical value by even a small amount, the system apparently becomes instable. In this model, the time step used was $0.01s$ to keep the system stable.

3 Brownian Dynamics (alternative procedure)

The global robustness of the network was studied by starting from a network of chemical reactions in noisy fluctuating environments:

$$\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x}) + \zeta \tag{9}$$

where $\mathbf{x} = (x_1(t), x_2(t))$is the concentration vector, with each component of which representing different protein species in the network. Then, $\mathbf{F}(\mathbf{x}) = (F_1(\mathbf{x}), F_2(\mathbf{x}))$ is the chemical reaction rate flux vector involving the chemical reactions which are often non-linear in protein concentrations $\mathbf{x}$ (for example, enzymatic reactions). The equations $\dot{\mathbf{x}} = \mathbf{F}(\mathbf{x})$ describe the averaged dynamical evolution of the chemical reaction network. Since within the cell the fluctuations of concentrations can be significant due to intrinsic and extrinsic noise [1] and in general can not be ignored, the noise term ζ is added for which Gaussian distribution is assumed (from the central limit theorem in statistics). Then the auto-correlations of noise is given by:

$$< \zeta(\mathbf{x}, t)\zeta^{\tau}(\mathbf{x}', t') >= 2D(\mathbf{x}, t)\delta(t - t') \; . \tag{10}$$

Here $\delta(t)$ is the Dirac delta function and the diffusion matrix D is explicitly defined by $< \zeta_i(t)\zeta_j(t') >= 2D_{ij}\delta(t - t')$. The average $< ... >$ is carried out with the Gaussian distribution for the noise.

We will follow the Brownian dynamics trajectories with multiple different initial conditions by solving the above stochastic differential equations itera-

tively [5]. We focus on the long time steady state properties and collect the statistics to obtain the steady state distribution function $P(\mathbf{x})$ for the state variable $\mathbf{x}$ (representing the protein concentrations of the cell fate network in this case). $P(\mathbf{x})$ is exponentially related to potential energy function $U(x)$:

$$P_0(\mathbf{x}) = \frac{1}{Z} \exp\{-U(\mathbf{x})/D\} , \tag{11}$$

with the partition function $Z = \int d\mathbf{x} \exp\{-U(\mathbf{x})/D\}$. From the steady-state distribution function, we can therefore identify $U(x)$ as the generalized potential energy function of the network system. In this way, we map out the "potential" energy landscape.

We calculated the Brownian dynamic trajectory of the movement using the Euler method. The length of each step was 0.3. The total steps are taken as 10^8 and the total 'jump' number for one parameter is more than 100. The two-dimensional space was then divided into lattice boxes which extended from the minimum value 0 to its maximum value 3 with steps of 0.01. The 10^8 points of data are put into the each lattice. The distribution $P(\mathbf{x})$ is then computed from the number of point inside each lattice box. The state of the lattice box with the most points is identified as the steady state for the system. The density distribution is the probability of each energy interval.

4 Transition Time, Barrier Height and Robustness

The stability of the network is related to the escape time from the basins of attraction. The easier it is to escape from the basins, the less stable of the network. For the probabilistic description of the network, the mean first-passage time for escape $\tau(x_1, x_2)$ starting from the point (x_1, x_2) obeys [6]: $F_1 \frac{\partial \tau}{\partial x_1} + F_2 \frac{\partial \tau}{\partial x_2} + D(\frac{\partial^2 \tau}{\partial x_1^2} + \frac{\partial^2 \tau}{\partial x_2^2}) = -1$. It is essentially the average time it takes from a initial position to reach a given final position. The equation can be

solved by an absorbing boundary condition at the given site and reflecting boundary conditions for the rest.

In Brownian Dynamics, the passage time was defined as the number of steps it takes the cell fate model to move(jump) from one state to another state. We calculated the mean passage time(MPT) τ during which the cell is in its S_C^*,S_A^* or S_B^* state.

The key point we want to make here in this paper is that due to the developmental changes, the system can have a preferential state to go to or direction of the flow for development by comparing the mean first passage transition times of forward and backward directions. This conclusion is based on the first passage time calculations, which by themselves are independent of whether the system is in equilibrium or non-equilibrium state because the forces that appeared in the corresponding first passage time equations do not have to be a gradient of a potential (details in the texts).

The first passage time has been addressed in [7, 8, 9, 10, 11] with minimum action, flux and other methods. Here we use two methods mentioned above which are different from those of the above-mentioned authors to calculate the mean first passage times. One is directly solving the partial differential equation that the first passage time satisfied with the appropriate boundary conditions. The other is through the Brownian dynamics simulations to check the results. Both methods agree with each other. The motivation here is that it is more likely that the developmental process is influenced more by the external fluctuations and hence, Fokker-Planck type of diffusion equations is suitable for the description (the number of the molecules are typically large so intrinsic fluctuations are relatively small). So our mean first passage transition time calculations are based on the fact that evolution of probability distribution obeys diffusion equation. Normally it is difficult to solve the N dimensional partial differential equations for mean first passage time this way. However, here we only consider two dimensions. So we can numerically exactly solve the corresponding partial differential equations for the mean

first passage time and compare with the Brownian dynamics simulations results.

For equilibrium dynamics, the transition rate theory holds which relates the barrier height with the transition first passage time. For non-equilibrium systems, the effect of the extra curl flux force is to deviate the kinetic paths from the gradient one. So the larger the curl flux force, the more deviations we expect from the transition state theory. We compared the two forces (the gradient and curl) in the relevant range of parameters we explore here. We found the gradient force is often larger than or perpendicular to the curl force in many regions of the state space. This means that in this parameter regime, the kinetic rate measured by the first passage time will be less influenced by the presence of the curl under those conditions. When we increase further the fluctuations so that it becomes large (- for this system we found that the critical value of D is 0.1 below which first passage time is correlated with the barrier height and above which there are less correlations.) , the landscape becomes flatter and the gradient force no longer dominates the curl. Then the first passage time is no longer controlled by the barrier height alone. It will be controlled by the total force with both gradient and curl.

5 Discussions

We want to discuss about landscape asymmetry towards unidirectional feature of the differentiation from multipotent state to differentiated state. It is correct that the specific direction of differentiation (which one of the (in this case) two available terminal attractor states will become occupied) obviously simply depends on the nature of the parameter change during fate commitment. This asymmetry induces change during differentiation depend on complex changes of environmental conditions leading to the changes of the landscape. This would correspond to "instructive" regulation, as opposed to stochastic regulation and is an old debate in stem cell biology. This refers

to the question of which one of the two available differentiated states the "arrow of time" points. However, here we do not address this problem explicitly. We only show that the directionality from the stem cell state attractors (central attractor) to either of the two marginal state is guaranteed because near the bifurcation that destabilizes the stem cell attractor, the "potential" of marginal (differentiated state) attractors are necessarily lower - as shown in the rest of the manuscript. Thus the discussion of the asymmetry in Fig.3 of main text is an example for how asymmetry in the landscape can be introduced to explain instructive regulation. The discussion was intended to show that the landscape picture here might provide a metaphor for understanding the process in a global way.

We wish to point out the advantages of the quantitative characterizations of the landscape for future stem cell research: (1). By characterization of the landscape topography of the stem cell, we can quantify the depth of the basins and barrier heights, which is crucial for marking the global state and stage in the development of stem cell. (2). We can also utilize the information on landscape topography to find out in which environmental and genetic condition would the stem cell undergo normal differentiation. (3). We can also utilize the information on landscape topography to find out in which environmental and genetic condition would the stem cell undergo reverse differentiation crucial for stem cell engineering. (4). We can use the quantification of the landscape to uncover the link between landscape topography and stability, suggest for design principle for stem cell engineering for function and robustness.

In concluding, what is new in this work are specific lessons learned from such analysis of global dynamics of particular circuits. It adds an additional dimension (potential) in the study of the properties of a system that is not acknowledged in the myriads of computational modeling studies of genetic circuits that are based on standard dynamical systems theory. These studies only present new "variations to a theme" by analyzing a specific (observed)

system structure but using standard tools.

In contrast, in this work, we apply a relatively novel idea of quasi-potential landscape to a particular circuit that drives cell fate decision. We show that by considering the global dynamics, epitomized by the potential landscape, a change in a control parameter is not only important with respect to bifurcations (existence of stable states) but has additional qualitative consequences: In our case it reveals that the directionality with which the bifurcation is traversed is relevant. The application of this robust phenomenon to stem cell differentiation explains the one-wayness of this process and thus, has biological implications not achieved by standard dynamical system analysis. Finally, we also would like to stress the importance of our specific analysis in that our 2-gene circuitry on which we apply the potential landscape approach has since the submission of our manuscript been popularized among stem cell biologists [12, 13]. However, these discussions do not provide a formal and quantitative explanations of the landscape.

6 Illustrations of existing models for directionality of differentiation

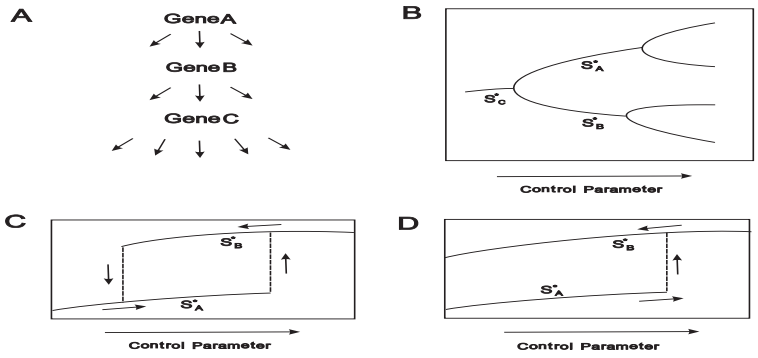

Figure 1: Existing models for directionality of differentiation. A. Signaling cascades acting akin to a "domino effect" sequentially activate or repress the sets of genes. B. Bifurcation diagram for a system that exhibits a sequential series of pitch-fork bifurcations as one control parameter (horizontal axis) changes in one direction. The solid lines ("branches") denote stable steady state values of the state variable x_1 (vertical axis). C. Hysteresis loop consisting of two stable branches (solid lines) representing the stable steady state values of the variable x_1. As the control parameter is altered the circuit "jumps" at critical parameter values along the dashed lines to the other branch. Note that jumping up and down between the two branches happens at different parameter values (see arrows) creating a hysteresis loop. D. Case of hysteresis in which due to alteration of system parameters (relative to the case in C.) other than that represented by the horizontal axis, the critical point for return to the lower branch disappears from the physically accessible state space.

References

[1] Kaern, M., T. C. Elston., W.J. Blake., and J. J. Collins. 2005. Stochasticity in gene expression: from theories to phenotypes. Nature Reviews 6: 451-464

[2] Kaufmann, B. B., and A. van Oudenaarden. 2007. Stochastic gene expression: from single molecules to the proteome. Curr. Opin. Genet. Dev. 17: 107-112.

[3] Wang, J., L. Xu, and E.K. Wang. 2008. Potential landscape and flux framework of nonequilibrium networks: robustness, dissipation, and coherence of biochemical oscillations. Proc. Natl. Acad. Sci. U. S. A. 105: 12271-12276.

[4] Hernjak, N., B. M. Slepchenko, K. Fernald, C.C. Fink, D. Fortin, I.I. Moraru, J. Watras, and L.M. Loew. 2005. Modeling and Analysis of Calcium Signaling Events Leading to Long-Term Depression in Cerebellar Purkinje Cells. Biophys. J. 89:3790-3806.

[5] McCammon, A., S. Harvey. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press. London.

[6] Risken H (1989) The Fokker-Planck Equation. Berlin (Springer).

[7] Maier, R.S., D.L. Stein. 1993. Escape problem for irreversible systems. Phys. Rev. E. 48, 931-938.

[8] Aurell, E., K. Sneppen.2002. Epigenetics as a first exit problem. Phys. Rev. Lett. 88, 048101-1.

[9] Walczak, A.M., J.N. Onuchic,P.G. Wolynes. 2005. Absolute rate theories of epigenetic stability. Proc. Natl. Acad. Sci. USA. 102, 18926-18931

[10] Roma, D.M., R.A. O'Flanagan, A.E. Ruckenstein, A.M. Sengupta, R. Mukhopadhyay. 2005. Optimal path to epigenetic switching. Phys. Rev. E. 71, 011902.

[11] Valeriani, C., R.J. Allen, M.J. Morelli, D. Frenkel, P.R.ten.Wolde. 2007. Computing stationary distributions in equilibrium and nonequilibrium systems with forward flux sampling. J Chem. Phys. 127:114109.

[12] Graf T., & T. Enver. 2009. Forcing cells to change lineages. Nature 462: 587.

[13] MacArthur, B.D., A. Ma'ayan, & I. R. Lemischka. 2009. Systems biology of stem cell fate and cellular reprogramming. Nat Rev Mol Cell Biol 10: 672.