Title
DNA-Binding Kinetics Determines the Mechanism of Noise-Induced Switching in Gene Networks

Journal
Biophysical Journal, 109(8)

ISSN
0006-3495

Authors
Tse, Margaret J
Chu, Brian K
Roy, Mahua
et al.

Publication Date
2015-10-01

DOI
10.1016/j.bpj.2015.08.035

Peer reviewed

Biophysical Journal Volume 109 October 2015 1746–1757

Article

DNA-Binding Kinetics Determines the Mechanism of Noise-Induced Switching in Gene Networks

Margaret J. Tse,[1] Brian K. Chu,[1] Mahua Roy,[1] and Elizabeth L. Read[1,2,*]
[1]Department of Chemical Engineering and Materials Science and [2]Department of Molecular Biology and Biochemistry, University of California Irvine, Irvine, California

ABSTRACT Gene regulatory networks are multistable dynamical systems in which attractor states represent cell phenotypes. Spontaneous, noise-induced transitions between these states are thought to underlie critical cellular processes, including cell developmental fate decisions, phenotypic plasticity in fluctuating environments, and carcinogenesis. As such, there is increasing interest in the development of theoretical and computational approaches that can shed light on the dynamics of these stochastic state transitions in multistable gene networks. We applied a numerical rare-event sampling algorithm to study transition paths of spontaneous noise-induced switching for a ubiquitous gene regulatory network motif, the bistable toggle switch, in which two mutually repressive genes compete for dominant expression. We find that the method can efficiently uncover detailed switching mechanisms that involve fluctuations both in occupancies of DNA regulatory sites and copy numbers of protein products. In addition, we show that the rate parameters governing binding and unbinding of regulatory proteins to DNA strongly influence the switching mechanism. In a regime of slow DNA-binding/unbinding kinetics, spontaneous switching occurs relatively frequently and is driven primarily by fluctuations in DNA-site occupancies. In contrast, in a regime of fast DNA-binding/unbinding kinetics, switching occurs rarely and is driven by fluctuations in levels of expressed protein. Our results demonstrate how spontaneous cell phenotype transitions involve collective behavior of both regulatory proteins and DNA. Computational approaches capable of simulating dynamics over many system variables are thus well suited to exploring dynamic mechanisms in gene networks.

INTRODUCTION

Multistable dynamics in gene regulatory networks has been proposed as the basis for the existence of diverse cell types (1–3). In this view, the biochemical interactions encoding gene networks give rise to complex, nonlinear expression dynamics. Distinct gene expression states—cellular phenotypes—are self-stabilizing attractors of the dynamical system. Transitions between attractors correspond to critical cellular processes, including developmental fate decisions (4,5), cellular reprogramming (3), phenotype switching (6), and carcinogenesis (7). As such, there is interest in characterizing the global dynamics of complex multistable gene networks to gain insight into the relative stability of cell states and the processes by which transitions between states can occur.

Over the past 20 years, the inherent noisiness of gene expression has been amply demonstrated. Stochastic biomolecular fluctuations can significantly impact dynamics of expression due to small-number effects (8–10). Positive feedback, which is common in regulatory networks, can amplify these fluctuations, giving rise to large-scale changes in gene expression programs. As such, intrinsic biomolecular noise is thought to underlie phenotypic variability in cell populations (11–14). Stochastic processes can also allow for spontaneous noise-induced transitions between metastable gene expression states. This stochastic state switching can be advantageous by priming cells to diversify according to alternative developmental programs (15,16) or by promoting survival of microorganisms or cancer cells in fluctuating environments (17–19). These findings highlight the need to understand how cellular networks achieve—or remain resistant to—noise-induced switching.

Common approaches to modeling gene networks formulate system dynamics as a set of stochastic biochemical reactions encompassing, for example, transcription-factor binding and unbinding to regulatory sites on DNA, gene transcription and translation, and degradation/dilution of mRNA and protein molecules. The associated master equations are amenable to Monte Carlo simulation (e.g., by the Gillespie algorithm (20)), which can exactly account for intrinsic stochastic fluctuations and molecular discreteness. Many studies of gene regulatory networks employ continuum approximations and dimensionality reduction to aid in the analysis of complex dynamics. For example, occupancies of binding sites on DNA are often assumed to be in quasiequilibrium with gene expression levels (i.e., concentrations of expressed proteins), which leads to nonlinear Hill function expressions for gene interactions (21–24). This assumption stems from a separation of timescales, when rates of binding and unbinding of regulatory proteins to DNA sites are fast relative to the rates of protein

Submitted May 19, 2015, and accepted for publication August 24, 2015.
*Correspondence: elread@uci.edu
Editor: Leah Edelstein-Keshet.

0006-3495/15/10/1746/12
http://dx.doi.org/10.1016/j.bpj.2015.08.035

synthesis and degradation. This kinetic regime has been termed the adiabatic limit, in analogy to electron transfer reactions (25).

Recent studies have focused on gene network dynamics in cases where this separation of timescales does not apply (4,26–29). In eukaryotic regulation, highly complex regulatory processes such as chromatin remodeling (in contrast to simple protein binding) drive fluctuations in gene expression. Such processes occur on relatively slow timescales; slow fluctuations in chromatin structure have been identified as a major source of gene expression noise in eukaryotes (9,10). Although knowledge of the detailed biochemical reactions underlying such higher-order regulation is generally lacking, theoretical studies have explored how different kinetic regimes contribute to the stability of gene expression states. Notably, cell-states have been found to be most stable (most resistant to noise-induced switching) in the adiabatic limit, where fast kinetics at DNA regulatory sites allows rapid response of gene expression states to local concentrations of regulatory proteins (28,30). This has led to the hypothesis that a weakly adiabatic kinetic regime allows for relatively frequent stochastic state switching, and thus developmental plasticity, in pluripotent stem cells (4,26,31).

In addition to exploring the principles governing the stability of gene expression states, recent theoretical studies have predicted transition paths (or reaction coordinates) of stochastic state switching in regulatory networks. These paths describe the most probable stepwise changes over multiple species in the network (e.g., the time-dependent changes in gene expression patterns) that occur as the system moves from one metastable state to another. Approaches for calculating transition paths have largely adopted analytical or numerical methods based on large deviation theory from chemical physics (32–36), related path integral approaches (4,5,27,37), or Monte Carlo simulations (31,38,39). Valuable insights into the dynamics of noise-induced switching in gene networks have emerged from these studies, including the significance of nonequilibrium phenomena (e.g., the irreversibility of switching paths) in gene network dynamics.

Gaining quantitative insight into stochastic state switching in gene networks is mathematically and computationally challenging. Stochastic fluctuations, overlapping temporal scales, and large numbers of interacting species can preclude the use of analytical approaches. Approximation methods may not accurately predict switching mechanisms driven by intrinsic biochemical fluctuations. For example, mean-field approximations have been found to distort multistable landscapes in some biochemical networks (40–43). On the other hand, brute-force Monte Carlo simulations can be plagued by inefficiency because state switching may be a rare event: long waiting times between transitions mean that simulations capture few, if any, switching events.

Numerical rare-event sampling algorithms provide an alternative approach by preferentially simulating events of interest without modifying the underlying system dynamics (44–46). Moreover, these algorithms aid interpretation of large amounts of noisy simulation data by providing automated means of extracting essential dynamical properties (47). These types of approaches have been adopted for the study of biochemical networks, including circadian rhythms (48), enzymatic cycles (46), and genetic switches (38). So-called string-based sampling algorithms (45,49,50) are especially well suited to discovery of transition paths for rare events in complex dynamical systems.

In this study, we employed a recently developed weighted-ensemble string (WE-string) simulation method (50) to study noise-induced switching in a set of genetic toggle-switch networks. By efficiently predicting transition paths involving collective changes over all species in the biochemical network, the approach uncovered new insights, to our knowledge, into how spontaneous switching depends on fluctuations of both protein number and DNA-binding occupancies. In particular, we found that the kinetics of protein binding and unbinding to DNA regulatory sites controls the mechanism of switching by determining whether the switching event is driven by these binding/unbinding events, or is instead driven by fluctuations in protein copy numbers. Our results potentially inform strategies to perturb gene network dynamics to stabilize or destabilize particular cellular states, or to drive a desired cellular transition.

MATERIALS AND METHODS

Modeling and simulation approach

Toggle-switch models

To explore the capability of the WE-string simulation method to predict transition paths in gene networks, we applied it to a set of related chemical network models for the genetic toggle switch. In this ubiquitous motif, two genes mutually repress one another through the action of their protein products, giving rise to a bistable gene expression pattern. Described in detail for the lysis/lysogeny decision of bacteriophage λ (51), the mutual repression motif is recognized as underlying a wide variety of cellular decisions, from embryonic development (52) to hematopoiesis (53) to specialization of T-cell subsets (54). The discovery of this motif paved the way for theory-based approaches to the design of synthetic gene circuits (23).

Quasipotential landscape

The stability of cell states can be expressed by a potential landscape over the gene network state space (5,55,56). We use a definition for a quasipotential function based on the stationary probability distribution: the quasipotential φ is taken to be $\varphi = -\ln(P_S)$, where P_S is the stationary (steady-state) probability over the state space (Fig. 1 *A*). Defined in this way, areas of state space with low potential are most probable, and therefore most stable. However, this quasipotential does not completely describe dynamics in nonequilibrium open systems such as biochemical networks (5,35,45).

Transition-path sampling algorithm

Predicting the most probable switching path by brute-force simulations is inefficient, because switching events are infrequent, and individual trajectories may be highly variable (Fig. 1 *A* and Figs. S1 and S2 in the Supporting Material). To predict the most probable path between two states

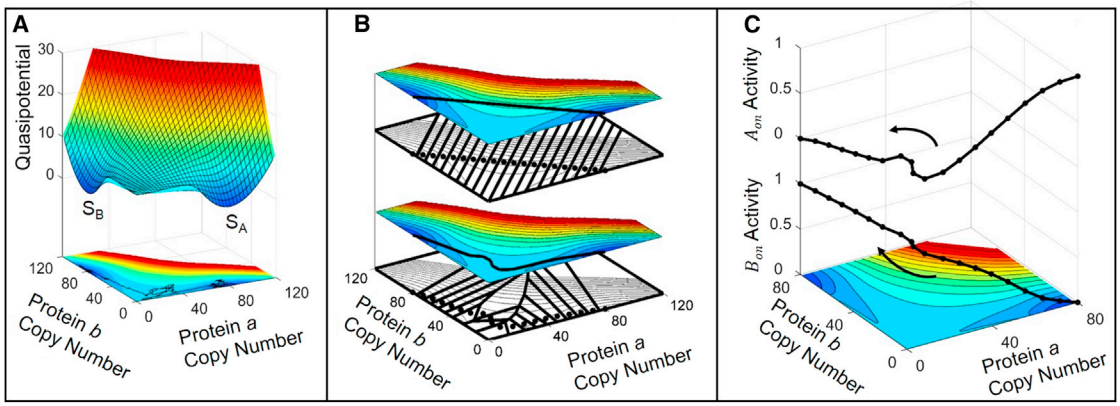

FIGURE 1 (*A*) (*Top*) The quasipotential surface, $\varphi = -\ln(P_S)$, for a bistable gene network. States with lower potential (*bluer colors*) are more stable (color scale applies to the whole figure). (*Bottom*) A single stochastic simulation trajectory that switches from the S_A attractor basin (high expression of protein *a* encoded by gene *A*) to the S_B basin (high expression of protein *b* by gene *B*). φ is projected onto subspaces representing *a* and *b* protein copy numbers. (*B*) The simulation method predicts the most probable transition path by utilizing a string connecting two states of interest to adaptively partition the state space and sample short stochastic trajectories initialized in each partition. (*Top*) The simulation is initialized with a string representing a guess path for the $S_A \rightarrow S_B$ transition (the corresponding Voronoi partitions are plotted underneath). (*Bottom*) After the simulation has converged, the string (and corresponding partitions) indicates the most probable transition path. (*C*) The predicted transition path contains information on all network species; for the toggle-switch network, it contains information on protein copy numbers and gene activities (DNA occupancies). (*Top*) Transition path for the $S_A \rightarrow S_B$ switch, projected onto *a*, *b*, and A_{on} subspaces, where A_{on} is the probability that the regulatory site for gene *A* is unbound, rendering it on (active). (*Bottom*) Same transition path projected onto *a*, *b*, and B_{on} subspaces, where B_{on} is the probability that the regulatory site for gene *B* is unbound.

of interest, the WE-string algorithm preferentially samples dynamics in the transition region (a low-probability area of state space), statistically merging information from many short trajectories. The algorithm achieves this by combining WE rare-event sampling (57) with a string-based adaptive discretization method for partitioning the state space (58). Adelman and Grabe recently introduced the method (50) and showed that it compares favorably to other path sampling methods in terms of efficiency, and could successfully sample dynamics in the space of many collective variables in an application to protein conformational change.

Rare-event sampling methods typically involve discretizing the state space of the system to access long-timescale kinetics on the basis of short-timescale trajectories in each region (59). The computational expense of discretizing a system evenly over *N* dimensions scales exponentially with *N*, which quickly becomes intractable for gene networks. The state space is naturally defined in terms of copy numbers and conformational/binding configurations of all biomolecular species, but even the smallest two-gene network (i.e., the toggle switch) has more than two species (including DNA promoter or regulatory binding sites, mRNA transcripts, protein products with multiple configurations, and so on, depending on the level of detail of the model). This challenge can be addressed by sampling along a single, relevant order parameter, such as a progress coordinate for the transition of interest, as pioneered by Allen and ten Wolde with the forward flux sampling (FFS) method (60). Alternatively, string-based methods adaptively partition space along a string: a one-dimensional path winding through a high-dimensional space, which connects two states of interest. This approach enables us to sample collective dynamics involving multiple system variables, without prior knowledge of an order parameter or the nature of dynamics in the transition region (61).

The string is defined by a set of N_{str} evenly spaced nodes, or points in the full state space; these nodes define the centers of N_{str} regions. All points in the space lie in a region that is defined by the nearest string node. Thus, the string nodes are the generating points of Voronoi polyhedra. The string is initialized as a guess path connecting two states (Fig. 1 *B*). For each iteration of the simulation, multiple weighted stochastic trajectories (replicas) in each region are simulated for a time τ. After each iteration, the WE method is used to statistically combine and duplicate weighted replicas as needed to ensure unbiased sampling of the partitioned space. After T_{move} iterations of τ, the string position is updated by moving the string nodes toward the average position of replicas in each region collected from the last T_{avg} iterations of τ, using a procedure of moving, smoothing, and reparameterization. Details of simulation parameters and algorithm implementation are in the Methods section. A more detailed description of the general algorithm is in Adelman and Grabe (50).

Because the string is defined as a list of points in the system state space, the converged string contains information about the probable stepwise changes along all network dimensions over the course of the transition. For our model toggle-switch network (Fig. 2), this means that the predicted transition path includes information about both gene expression levels (copy numbers of protein products *a* and *b*) and gene activity states (i.e., likelihood of gene states A_{on} and B_{on}), which reflect the occupancy states of the DNA-binding sites (Fig. 1 *C*).

Methods

Toggle-switch reaction network

We considered two previously studied variants of the toggle switch, which we term the base network (40) and the network with explicit dimerization (Fig. 2 and Eqs. 1–3) (38). In both variants, protein dimers act as repressors of the other competitor gene. The switch is symmetric with respect to the rate parameters governing the behavior of each of the two genes, *A* and *B*. Translation and transcription are subsumed into a single biochemical reaction for synthesis of the two encoded proteins, *a* and *b*. The two genes exist in one of two occupancy states, giving rise to binary regulation ($A_{on/off}$, $B_{on/off}$), where only the on states contribute to protein synthesis. We furthermore studied the general and exclusive versions of the switch; the exclusive switch assumes that the a_2 and b_2 repressors competitively bind to a single regulatory site that controls both genes; therefore, both genes cannot simultaneously be in the off state (i.e., the doubly repressed state is disallowed (equations are in the Supporting Material)).

Protein synthesis and degradation reactions are given by

$$\begin{array}{cc} A_{on} \xrightarrow{g} A_{on} + a, & B_{on} \xrightarrow{g} B_{on} + b \\ a \xrightarrow{k} \varnothing, & b \xrightarrow{k} \varnothing \end{array}. \qquad (1)$$

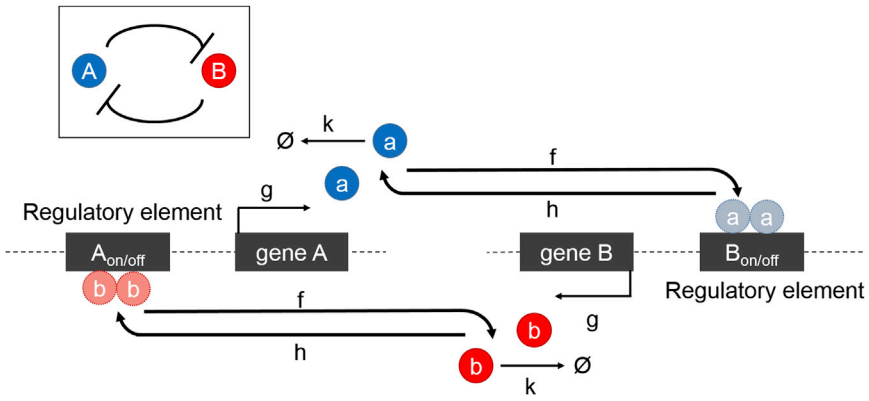

FIGURE 2 Schematic diagram of biochemical reactions in the genetic toggle switch network (the inset shows the basic network motif of two mutually repressing genes). Gene A codes for repressor protein a (*blue*) and gene B codes for protein b (*red*). Each gene is controlled by a regulatory element (e.g., the promoter). When the regulatory element is unbound, the gene is on (active), or free to express protein. When it is bound by a homodimer of the opposing repressor, the gene is off (inactive). To see this figure in color, go online.

In the base network, dimerization is neglected, and gene repression is represented by

$$A_{\mathrm{on}} + 2b \underset{f}{\overset{h}{\rightleftharpoons}} A_{\mathrm{off}}, \quad B_{\mathrm{on}} + 2a \underset{f}{\overset{h}{\rightleftharpoons}} B_{\mathrm{off}}. \tag{2}$$

These reactions give rise to an overall rate of gene inactivation of $hx(x-1)/2$, where x is the repressor protein encoded by the competing gene. In the network with explicit dimerization, the above reactions are replaced by

$$\begin{aligned} A_{\mathrm{on}} + b_2 &\underset{f}{\overset{h}{\rightleftharpoons}} A_{\mathrm{off}}, \quad B_{\mathrm{on}} + a_2 \underset{f}{\overset{h}{\rightleftharpoons}} B_{\mathrm{off}}, \\ 2a &\underset{u}{\overset{d}{\rightleftharpoons}} a_2, \qquad\qquad 2b \underset{u}{\overset{d}{\rightleftharpoons}} b_2 \end{aligned} \tag{3}$$

Rate parameters for the reaction network are given in Table 1. The effect of DNA-binding kinetics was studied by varying the rates h and f (binding and unbinding, respectively, of transcription factors to DNA). The binding equilibrium constant $X_{\mathrm{eq}} = f/h$ was maintained at a constant value, such that the locations of the two stable states are preserved.

Calculation of quasipotential

We calculated P_{S} and φ by expressing the chemical master equation (CME) for the network in matrix form: $(dP/dt) = AP$, where A is the transition-rate matrix over a truncated state space $(0 \le a, b \le 120)$, and solving for the eigenvector associated with the zero eigenvalue (62) using MATLAB (63). For parameter regimes giving relatively short waiting times between switching events (i.e., parameter sets II and III), we also obtained P_{S} from a single long Gillespie simulation trajectory, using BioNetGen (version 2.2.2) (64). The quasipotential landscapes obtained from each method were indistinguishable.

Transition-path simulation

The WE-string sampling method was adapted from Adelman et al. (50). The procedure for string smoothing, parameterization, and separate sampling of forward and backward paths (by separating replicas according to their most recently visited basin) was adapted from Dickson et al. (48). String ends were fixed in the two stable attractor basins, S_A and S_B, corresponding to states with high gene A expression and high gene B expression, respectively. S_A and S_B were defined as hyperspheres with unit radius in N-dimensional state space (base network, $N = 4$ {a, b, $A_{\mathrm{on/off}}$, $B_{\mathrm{on/off}}$}; explicit-dimerization network, $N = 6$ {a, b, a_2, b_2, $A_{\mathrm{on/off}}$, and $B_{\mathrm{on/off}}$ }). The basin centers were found by identifying the minima of the quasipotential for each parameter set. The number of string nodes, N_{str}, was 20, and the number of replicas in each region (N_{rep}) was 150. Definitions and values of additional simulation parameters are given in Table S1. Switch progress is defined as the normalized distance along the converged string. That is, for the forward path, the progress at the ith string node is defined by D_i/D_{AB}, where D_i is the Euclidean distance to the ith node from the center of S_A, and D_{AB} is the total distance along the string. The simulation was determined to have converged after the position of the string remained stable over >1000 string movements. Convergence was determined from the

TABLE 1 Rate parameters and calculated $S_A \rightarrow S_B$ switching rates, k_{AB}, for all studied genetic toggle-switch variants

					k_{AB} [t^{-1}] (General)		k_{AB} [t^{-1}] (Exclusive)	
Network Parameters					Mean	Std.	Mean	Std.
Base Network								
	h [k^{-1}]	f [k^{-1}]	g [k^{-1}]	k [t^{-1}]				
I	10^2	10^4	80	1	1.9×10^{-6}	2×10^{-7}	5.8×10^{-10}	6×10^{-11}
II	10^{-1}	10	80	1	3×10^{-4}	3×10^{-4}	3×10^{-4}	3×10^{-4}
III	10^{-2}	1	80	1	7×10^{-3}	7×10^{-3}	3×10^{-3}	3×10^{-3}
Explicit Dimerization								
	h [u^{-1}]	f [u^{-1}]	g [u^{-1}]	k [u^{-1}]	d [u^{-1}]	u [t^{-1}]		
I	8.89×10^3	10^5	80	8.944	1	1	3.3×10^{-10}	7×10^{-11}

Included are parameters for base and explicit-dimerization networks for both general and exclusive switch versions. Rate parameters correspond to the schematic in Fig. 2 (see Methods): h and f, binding and unbinding rates, respectively, of proteins to DNA; g, protein synthesis; k, protein degradation; d and u, binding and unbinding rates, respectively, of monomer proteins to form homodimers. All network parameters are reported in units of [k^{-1}] (the inverse degradation rate) in the base network and [u^{-1}] (the inverse dimer unbinding rate) in the explicit-dimerization network. Parameter set I corresponds to the adiabatic regime, where the propensities of protein binding to and unbinding from DNA ($hx(x-1)/2$ and f, respectively) are large relative to g and k. Std., standard deviation.

root mean-squared difference between the current node positions and the running average of the previous 100 positions (Fig. S3). Sampling code was written in MATLAB and Gillespie trajectories were simulated using BioNetGen (64).

Calculation of switching rates

Switching-rate constants, k_{AB}, were estimated from brute-force Gillespie simulations when possible (the faster switching regime: parameter sets II and III) or WE sampling in cases where capturing switching events from brute-force simulations was intractable (the rare-switching regime: parameter set I). The network parameters are symmetric, such that $k_{AB} = k_{BA}$. For the brute-force simulations, k_{AB} was estimated by inverting the mean first-passage time of $(S_A \rightarrow S_B)$ transitions from 1000 trajectories. The WE rate estimation followed (65), using a linear partitioning of state space along the progress coordinate λ with bins of unit length. λ is given by $\lambda = na - nb$, where $na = a + 2B_{\text{off}}$, $nb = b + 2A_{\text{off}}$ (general switch) or $na = a + 2a_2 + 2B_{\text{off}}$, $nb = b + 2b_2 + 2A_{\text{off}}$ (exclusive switch); this is the same progress coordinate as used previously (38,60). The simulation time step was chosen to be the same as that of the WE-string simulation for the corresponding parameter set; 150 replicas were simulated in each bin.

Validation of transition-path simulation

Validation of switching mechanisms predicted by the transition-path simulations was done by comparison to brute-force switching trajectories (where possible), and by committor analysis (66). Successful switching trajectories harvested from brute-force simulations were binned according to the progress coordinate λ. Additional simulations assessed committor probabilities along the converged strings: these reflect the probability that simulation replicas in a given region will next commit to the final target state rather than returning to their most recently visited basin.

RESULTS

Stability of phenotypes depends on DNA-binding kinetics

We calculated the quasipotential landscape for the genetic switch, and varied the parameters governing DNA-binding/unbinding kinetics between the adiabatic (fast kinetics (Figs. 3 and 4, parameter set I)) and nonadiabatic (slow kinetics (Figs. 3 and 4, parameter set III)) regimes. By varying rates of unbinding (f) and binding (h) while maintaining their constant ratio, $X_{\text{eq}} = f/h$, the bistability and locations of the attractor basin centers are preserved, but the barrier height separating the two states changes (Figs. S4 and S5). Our results show that the barrier height decreases with decreasing adiabaticity, in agreement with previous results (40). This difference is also reflected in the calculated rate constants, k_{AB}, for spontaneous switching from the attractor state S_A to S_B (Table 1). The rate constant for spontaneous switching between metastable states has been identified as a measure of stability or robustness of gene expression states to noise. In agreement with previous studies (4,26,28,30,40), we find that the rate of switching slows with increasing adiabaticity, indicating that gene expression states are more stable with faster DNA-binding kinetics. This trend was preserved for all variants of the switch (Table 1). For parameter set I, k_{AB} is much lower for the exclusive than for the general switch, whereas the switching rates for the two variants are of the same order of magnitude for sets II and III. Interestingly, the switching rates are nearly identical for parameter set II.

Sampling algorithm finds a transition path for the toggle switch

Transition paths of noise-induced switching between gene expression states were predicted by performing WE-string sampling of stochastic simulations. Transition paths were predicted for both forward $(S_A \rightarrow S_B)$ and backward $(S_B \rightarrow S_A)$ switching events and were compared to the numerically calculated quasipotential landscapes for the three parameter sets. Although the WE-string simulations require no foreknowledge of the underlying potential in the transition region, the paths converge to the relatively low-potential (i.e., high-probability density) transition tube connecting the two states (Figs. 3 and 4). In addition to mapping out the region of state space traversed by successful transitions, committor analysis (Figs. S9 and S10 and Methods) indicates that for most studied parameter sets, the string also accurately reflects the true reaction coordinate for the transition. Progress along the string corresponds to the most probable step-by-step changes undergone as the system progresses from one basin to another. However, slight discrepancies exist for the general switch in the nonadiabatic regime, likely due to a known limitation of string-based sampling methods in cases of wide transition tubes (58). Predicted transition paths were robust to differences in simulation initial conditions (guess paths) (Fig. S6).

Transition paths reveal additional dynamic features and nonequilibrium phenomena

The predicted transition paths also reveal that the system dynamics is not governed solely by the topography of the potential surface. The paths reveal additional features, including pathway oscillations and nonoverlapping forward and reverse paths, in agreement with previous studies (5,38). These features can be seen by comparison to calculated minimum energy paths (Figs. S7 and S8), which reflect only gradient dynamics over the projected quasipotential surface. The degree to which the forward and reverse paths diverge depends both on the parameters and on the choice of subspace on which the system dynamics is projected. For example, when projected onto the protein copy number subspace, the paths show more divergence in the nonadiabatic regime than in the adiabatic regime for the general switch (Fig. 3), whereas the opposite is true for the exclusive switch (Fig. 4). For the exclusive switch in the nonadiabatic regime (Fig. 4, parameter set III), there is clear divergence of the forward and reverse transition paths in the subspace of DNA-occupancy states (gene states), yet the transition paths are nearly superimposed in the protein subspace (Fig. 4.).

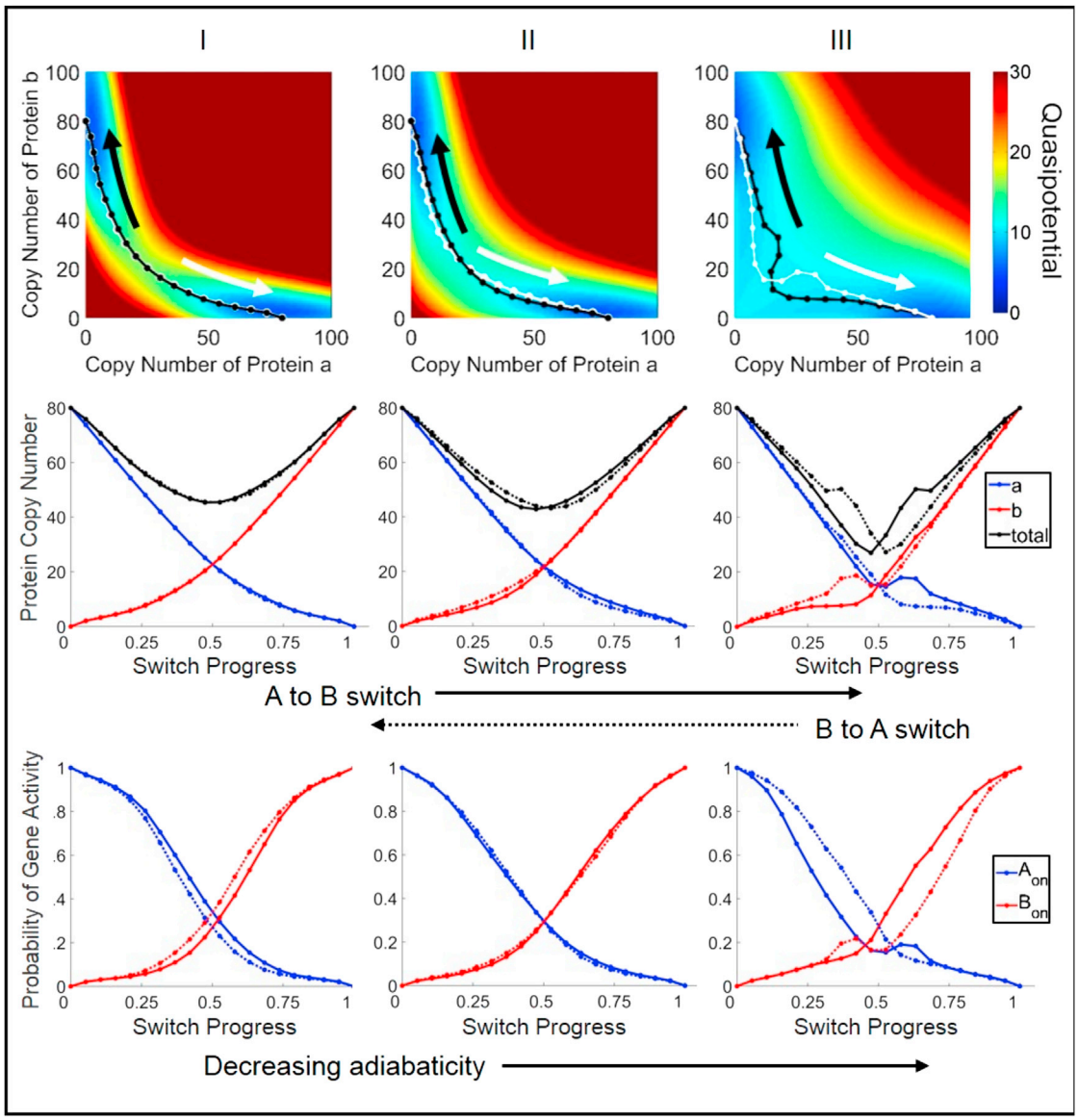

FIGURE 3 Quasipotential surfaces and predicted transition paths for the base general-toggle-switch network. (*Left* to *right*) Network rate parameters with decreasing adiabaticity (slower DNA-binding kinetics) (see Table 1 for parameters). (*Top*). Transition paths are superimposed on the two-dimensional projection of the quasipotential surface. The black and white paths illustrate the forward ($S_A \rightarrow S_B$) and backward ($S_B \rightarrow S_A$) transitions, respectively. (*Middle*) Transition paths are plotted as protein copy numbers versus switch progress (protein *a*, *blue*; protein *b*, *red*; total protein number ($a + b$), *black*). Switch progress is defined as the normalized distance along the transition path (see Methods). Forward switching ($S_A \rightarrow S_B$) is displayed as solid lines and backward ($S_B \rightarrow S_A$) as dotted lines. (*Bottom*) Transition paths are plotted as gene activities versus switch progress. The activity of gene *A* (A_{on}, *blue curve*) is the probability of the regulatory site of gene *A* being unbound, which renders it active. This is similar for B_{on} (*red curve*).

Protein number oscillations are observed in the adiabatic regime for the exclusive switch (Fig. 4, parameter set I). This motion is not echoed in the DNA binding occupancies, suggesting that it results solely from the birth-death reactions for protein expression. In the nonadiabatic regime for the general switch (Fig. 3, parameter set III), oscillatory motion in the proteins is echoed in the DNA-binding occupancies (gene activities), suggesting that it could result from previously described eddy currents in gene networks, where slow transcription factor binding/unbinding events drive cyclic dynamics of protein expression in the nonadiabatic limit (26,30,37). However, as discussed above, the detailed features of the string for the nonadiabatic general switch may not reflect dynamics along the true reaction coordinate.

Influence of DNA-binding kinetics on the switching mechanism

The predicted transition paths for the toggle switch reveal that the switching mechanism is altered by the DNA-binding

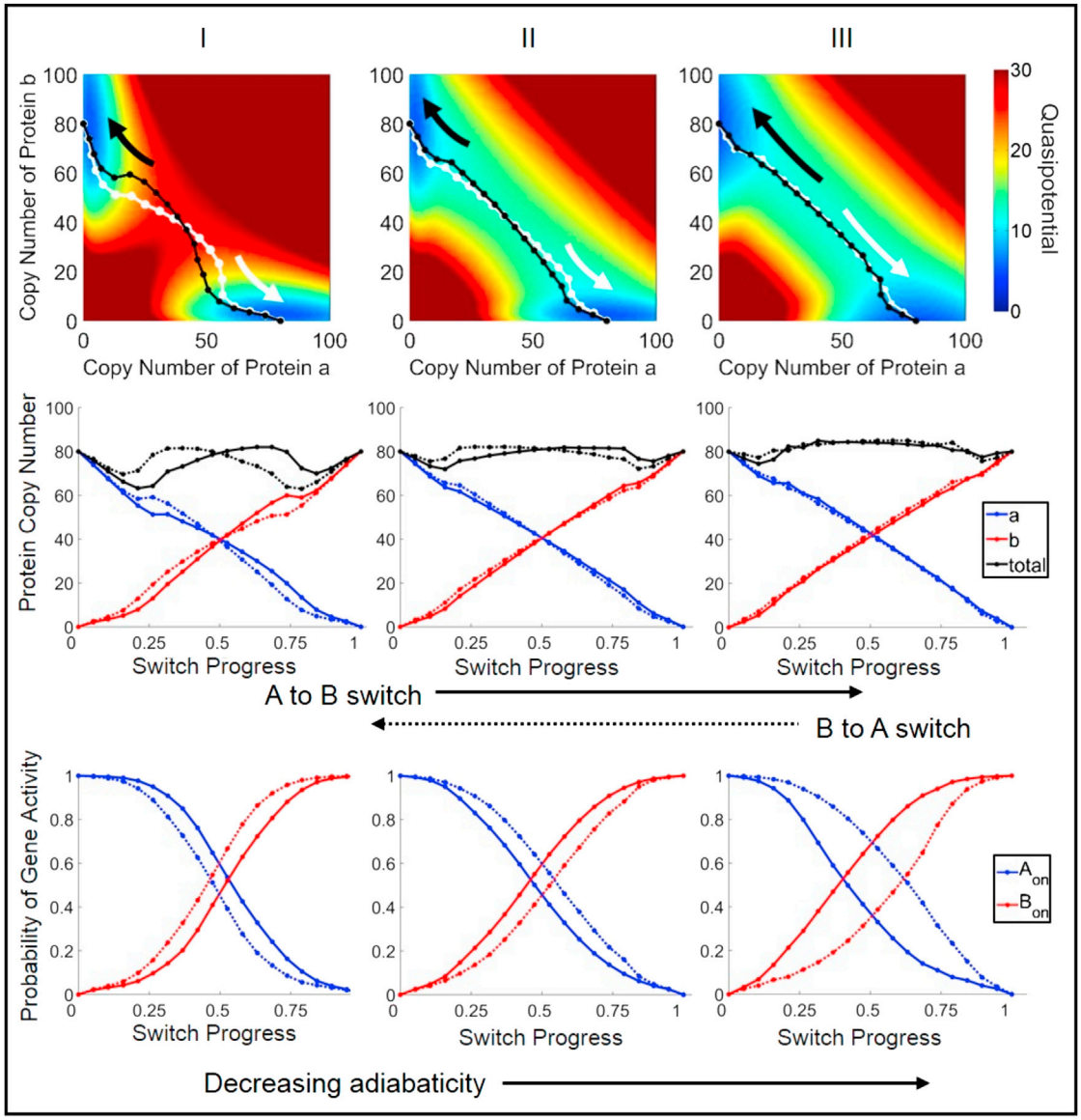

FIGURE 4 Quasipotential surfaces and predicted transition paths for the exclusive-toggle-switch network, with competitive binding of repressors to regulatory sites. All definitions are the same as in Fig. 3. (Left to right) Network rate parameters with decreasing adiabaticity (slower DNA-binding kinetics) (see Table 1 for parameters). (*Top*). Transition paths are superimposed on the two-dimensional projection of the quasipotential surface. (*Middle*) Transition paths are plotted as protein copy numbers versus switch progress. (*Bottom*) Transition paths are plotted as gene activities versus switch progress.

kinetic parameters. This is seen by considering the switch progress separately in terms of either protein copy numbers or gene activities: the switch progress in these subspaces is not synchronized, and the degree to which they are asynchronous depends on the DNA-binding kinetics. In this symmetric system, when the system reaches the separatrix at ($a = b$), the transition can be considered to be half-complete. This crossing corresponds closely to the half-distance point along the transition path (switch progress = 0.5) (Figs. 3 and 4). Under a quasiequilibrium assumption, the gene activities depend directly on the levels of protein expression according to $A_{on} = 1/(1 + b^2/X_{eq})$ and $B_{on} = 1/(1 + a^2/X_{eq})$. As such, when protein expression from the two competing genes reaches equivalence, the genes would then also have equal probabilities of being in the active (unbound) states, giving ($A_{on} = B_{on}$). However, the transition paths reveal that this crossing is not necessarily reached simultaneously in both subspaces. Similar trends are seen in the parameter dependence for the different switch variants: in the adiabatic regime, the gene activities reach ($A_{on} = B_{on}$) after the protein numbers reach ($a = b$) (Figs. 3 and 4, *left column*), whereas in the nonadiabatic regime, the gene activities equalize before the proteins (Figs. 3 and 4, *right column*). In other words, when DNA-binding kinetics is slow, stochastic binding and unbinding events play a larger role in driving the switch toward completion, and protein

expression follows. Conversely, when DNA-binding kinetics is fast, birth-death fluctuations in protein numbers drive the switch. This trend was preserved for additional studied parameter sets (Figs. S11 and S12) and also appeared in averaging of switching trajectories harvested from brute-force simulations (Figs. S1 and S2).

The simulations predict that the exclusive and general switch variants (see Methods) progress through different transition states. Our results are in agreement with the previous finding (38) that the general switch transitions through a state in which both genes are repressed, which is also reflected in the decrease in overall expression of proteins (Fig. 3, total expression curves). For the exclusive switch variant, the transition progresses through a region where each gene has a nearly equal probability of being on or off. Additionally, the total protein number when the switch is half-complete (switch progress = 0.5) remains high after an initial decrease (Fig. 4, *total*). These transition states clearly reflect the difference in the DNA occupancy states available to each toggle-switch variant. In the general switch, all combinations of bound and unbound states at the regulatory sites of the two genes are allowed (i.e., A_{on}/B_{on}, A_{on}/B_{on}, A_{off}/B_{on}, A_{off}/B_{off}), whereas in the exclusive switch, the A_{off}/B_{off} state is not accessible due to competitive binding in the case of overlapping regulatory sequences. This scenario is the limiting case of a reduced binding rate of a repressor if the competing repressor is already bound to the DNA.

Advantage of the transition-path simulation method for multidimensional networks with rare switching

We applied the sampling algorithm to the extended network with explicit dimerization (Methods). The parameters of the extended network were chosen such that the base and extended networks give identical total protein expression at steady state, in terms of the variables *na* and *nb* (see Methods and the Supporting Material). The predicted transition paths showed qualitatively similar switching mechanisms for the base and extended networks. In particular, the trend of asynchronous switching in protein numbers and gene activities was preserved in the network with explicit dimerization, with gene activities lagging behind protein numbers in the adiabatic regime (Fig. 5). For this parameter set, it was not possible to access the full quasipotential surface: because the size of the state space increases exponentially with dimensionality (i.e., number of species), the addition of two more molecular species (the dimers a_2 and b_2) renders the transition rate matrix too large for straightforward numerical calculations (see Methods). This curse of dimensionality does not limit Gillespie simulations; however, brute-force simulation of 10^9 time steps failed to capture a single switching event (Fig. 5 *A*) due to the low probability of switching ($k_{AB} = 3.3 \times 10^{-10}$, Table 1). Our results thus demonstrate that the WE-string sampling method is particularly advantageous for studying transition paths for networks with several or more molecular species and long waiting times between events of interest.

DISCUSSION

In this work, we applied recently developed stochastic simulation methods to the study of gene regulation dynamics to predict switching mechanisms in a ubiquitous bistable gene network. Our simulations recapitulate several general aspects of gene network switching, including the increased stability of cell states to noise-induced switching with fast DNA-binding/unbinding kinetics and evidence of nonequilibrium phenomena in switching dynamics, in agreement with previous studies (4,5,28,30,38,40). Our simulations also uncovered new insights, to our knowledge, into stochastic phenotype switching. In particular, we found that switching progresses asynchronously in the separate

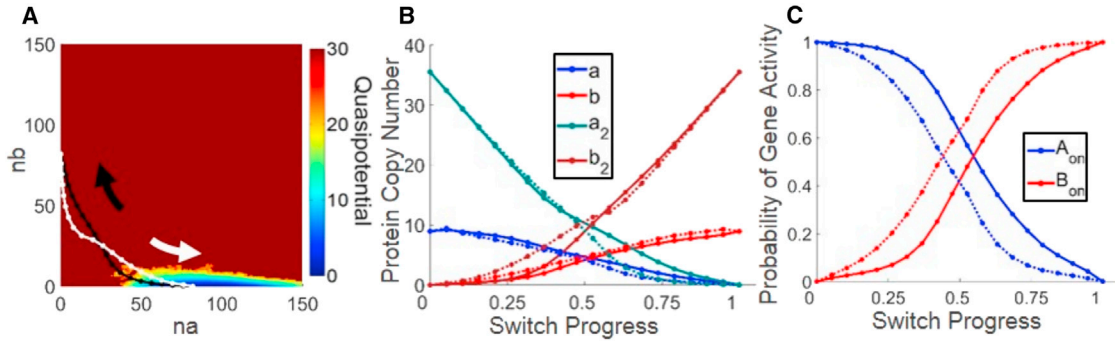

FIGURE 5 Transition paths predicted for the exclusive switch with explicit dimerization, parameter set I (see Table 1), and incomplete quasipotential surface. (*A*) Predicted transition paths for the $S_A \rightarrow S_B$ (*black*) and $S_B \rightarrow S_A$ (*white*) transitions. The two-dimensional projection shows the quasipotential surface obtained from a brute-force Gillespie simulation of 10^9 time steps initialized in S_A. The long simulation fails to produce a single switching event, giving rise to a one-sided, incomplete quasipotential surface. Despite the rarity of switching for this network ($k_{AB} = 3.3 \times 10^{-10}$ (Table 1)), the simulation method remains capable of predicting the most probable transition path. (*B*) Transition paths are plotted as protein copy numbers versus switch progress. (*C*) Transition paths are plotted as gene activities versus switch progress.

subspaces of protein copy numbers or DNA occupancies (i.e., gene activities). We observed a trend wherein the gene activities lagged behind the protein numbers during the switching transition in the fast-DNA-binding (adiabatic) kinetic regime and switched ahead of the proteins in the slow-DNA-binding (nonadiabatic) regime. Thus, our results reveal that the detailed mechanisms (and overall rates, k_{AB}) of spontaneous switching depend on the network parameters. Our findings suggest that protein copy-number fluctuations play a more significant role than DNA-site-occupancy fluctuations in driving the switch toward completion in the adiabatic regime, whereas the converse is true in the nonadiabatic regime. Our study also demonstrates that rare-event sampling algorithms, in particular string-based methods (49,50), are well-suited to the discovery of transition paths that link metastable states in gene networks.

Stochastic state switching in gene networks is thought to be a mechanism driving phenotypic heterogeneity in genetically identical cell populations. This type of nongenetic heterogeneity induced by transient switching events appears to play a role in diverse biological contexts, including development and disease. In stem cell networks, probabilistic differentiation has been linked to stochastic switching of pluripotent cells between substates, distinguished by different levels of expression of key transcription factors (67–70). Stochastic state switching is also thought to allow bacteria to switch into and out of antibiotic-resistant states (17), virus-infected cells to switch into and out of latency (12), and cancer cells to switch into and out of chemotherapy-resistant states (19). Identification of the mechanisms by which these switching events occur can provide a route to discovery of novel strategies for cellular engineering (e.g., stem cell reprogramming) and drug treatment. However, mechanistic studies of switching can be experimentally challenging, particularly when only a fraction of cells in a population is poised to transition between states (71). As such, quantitative models coupled with the computational approaches presented here could provide important insight into underlying mechanisms, e.g., by predicting patterns of gene expression associated with cells that are in transit between phenotypic states.

Chromatin remodeling contributes to gene expression noise by stochastically transitioning individual genes between on (active) and off (inactive) states (9,10). These transitions can occur on timescales that are of the same order as or longer than mRNA and protein lifetimes (72). Though the toggle-switch model studied here is highly simplified, the slow-DNA-binding (weakly adiabatic or nonadiabatic) kinetic regime broadly captures this complex epigenetic regulation. The relative instability of phenotypic states in this regime has been hypothesized to underlie developmental plasticity of pluripotent stem cells (4,26). Our results suggest that stochastic fluctuations at the level of the individual promoter sites, rather than changes in local regulatory protein concentration, play a critical role in driving the spontaneous, global switching events that poise pluripotent cells to choose probabilistically between alternative lineages.

The simulation method employed in this study simultaneously tackles two challenges encountered in modeling gene network dynamics: the rare-event problem (switching between gene states may occur rarely, giving rise to impossibly long simulation convergence times) and the curse of dimensionality (switching requires coordinated changes involving many species in the network). By applying the WE-string method to different toggle-switch variants in multiple kinetic regimes, we demonstrate its potential flexibility for capturing stochastic dynamics in different gene network systems. We found that a major advantage of the method is its ability to simultaneously track dynamics of both protein products and DNA occupancy states (i.e., promoter states), which together control global gene network dynamics.

In previous studies of transition paths in gene networks where DNA occupancies were treated as being in quasiequilibrium with protein numbers, it was implicitly assumed that the dynamics in the two network subspaces are identical (32–36).Our results show that even in the adiabatic regime, where a separation-of-timescales assumption is justified, the switching dynamics of the gene activities and protein numbers have distinct features. Theoretical methods have been developed recently for studying noise-induced transitions in gene networks, which can account for nonadiabaticity (slow DNA binding) by path integral approaches based on approximations to the CME (4,37). In this study, we explore an alternative approach in direct sampling, which circumvents the need for approximations to the CME, can be used in conjunction with available stochastic simulation software packages (64), and potentially scales more easily to other types of biochemical networks with many species.

The application of rare-event sampling techniques to study switching dynamics in gene networks was pioneered previously with the FFS method (60). Morelli et al. (38) carried out a detailed study of the toggle switch by FFS; our network model is largely based on theirs, although we studied parameter regimes commensurate with those of Sasai et al. (25), producing larger average protein numbers. The transition states we found for the general versus exclusive switch variants are in qualitative agreement with that study (38), although direct quantitative comparison of our results to theirs is not possible because of the different parameter regimes and because their study did not focus explicitly on the mechanistic dependence on adiabaticity.

In principle, FFS and string-based sampling methods can access the same information; the advantages and disadvantages of each method, making them potentially complementary, have been discussed previously (61). FFS samples dynamics along a single order parameter, which must be defined a priori, and which tracks progress between initial

and final states. The choice of this parameter is nontrivial, and a poor choice can lead to computational inefficiency, although statistical methods to optimize the order parameter have been developed (73). Use of a single order parameter necessarily results in loss of information about dynamics along separate system coordinates; the progress coordinate λ is much more sensitive to protein copy numbers than to DNA-occupancy states. Adaptive string-based methods were developed to circumvent this problem and access collective dynamics in spaces of many order parameters (45). Because progress through a transition is measured by successive string nodes, each of which represents a configuration in the full state space, this approach offers a more intuitive description of dynamics in gene networks with many species. Moreover, the approach enabled us to directly isolate separate contributions of protein numbers and DNA-occupancy states to the switching mechanism, which in turn revealed the mechanistic dependence on adiabaticity.

However, string methods also have potential limitations. Convergence of the string to the most probable transition path may not occur in systems with highly complex quasi-potential landscapes with several or more local minima, so extension of the method to more complex gene networks remains to be explored. The transition path predicted by the converged string may not correspond exactly to the true chemical reaction coordinate for some networks (as defined by committor probabilities (66)) due to the difficulty of accurately partitioning wide transition tubes (58). Our results show that despite these potential limitations, the WE-string method was able to resolve switching dynamics over a wide range of parameters. Even in the nonadiabatic limit, the major mechanistic predictions (regarding the nature of the transition state and the asynchronous switch progression in the protein versus gene activity subspaces) were found to be robust by comparison to brute force trajectories. Our findings suggest that the method may be particularly useful for exploring dynamics of gene networks where little information (such as a suitable progress coordinate) is available a priori and may prove particularly powerful in combination with other, complementary rare-event sampling approaches.

The code used to generate the data in this article will be accessible from https://github.com/Read-lab/WE-string-GRN.git.

SUPPORTING MATERIAL

Supporting Materials and Methods, twelve figures, and one table are available at http://www.biophysj.org/biophysj/supplemental/S0006-3495(15)00870-X.

ACKNOWLEDGMENTS

We thank the administrators of the University of California, Irvine High-Performance Computing cluster.

SUPPORTING CITATIONS

Reference (74) appears in the Supporting Material.

REFERENCES

1. Kauffman, S. 1973. Control circuits for determination and transdetermination. *Science.* 181:310–318.
2. Huang, S., G. Eichler, …, D. Ingber. 2005. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys. Rev. Lett.* 94:128701.
3. Huang, S. 2009. Reprogramming cell fates: reconciling rarity with robustness. *BioEssays.* 31:546–560.
4. Zhang, B., and P. Wolynes. 2014. Stem cell differentiation as a many-body problem. *Proc. Natl. Acad. Sci. USA.* 111:10185–10190.
5. Wang, J., K. Zhang, …, E. Wang. 2011. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc. Natl. Acad. Sci. USA.* 108:8257–8262.
6. Hong, T., J. Xing, …, J. Tyson. 2011. A mathematical model for the reciprocal differentiation of T helper 17 cells and induced regulatory T cells. *PLOS Comput. Biol.* 7:e1002122.
7. Huang, S., and S. Kauffman. 2013. How to escape the cancer attractor: Rationale and limitations of multi-target drugs. *Semin. Cancer Biol.* 23:270–278.
8. Beard, D. A., and H. Qian. 2008. Chemical Biophysics: Quantitative Analysis of Cellular Systems. Cambridge University Press, Cambridge, United Kingdom.
9. Raj, A., and A. van Oudenaarden. 2008. Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell.* 135:216–226.
10. Miller-Jensen, K., S. Dey, …, A. Arkin. 2011. Varying virulence: epigenetic control of expression noise and disease processes. *Trends Biotechnol.* 29:517–525.
11. Arkin, A., J. Ross, and H. McAdams. 1997. Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics.* 149:1633–1648.
12. Weinberger, L. S., J. C. Burnett, …, D. V. Schaffer. 2005. Stochastic gene expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity. *Cell.* 122:169–182.
13. Hume, D. A. 2000. Probability in transcriptional regulation and its implications for leukocyte differentiation and inducible gene expression. *Blood.* 96:2323–2328.
14. Choi, P. J., L. Cai, …, X. S. Xie. 2008. A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science.* 322:442–446.
15. Chang, H. H., M. Hemberg, …, S. Huang. 2008. Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature.* 453:544–547.
16. Dietrich, J.-E., and T. Hiiragi. 2007. Stochastic patterning in the mouse pre-implantation embryo. *Development.* 134:4219–4231.
17. Balaban, N., J. Merrin, …, S. Leibler. 2004. Bacterial persistence as a phenotypic switch. *Science.* 305:1622–1625.
18. Acar, M., J. T. Mettetal, and A. van Oudenaarden. 2008. Stochastic switching as a survival strategy in fluctuating environments. *Nat. Genet.* 40:471–475.
19. Sharma, S., D. Lee, …, J. Settleman. 2010. A chromatin-mediated reversible drug-tolerant state in cancer cell subpopulations. *Cell.* 141:69–80.
20. Gillespie, D. 1977. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* 81:2340–2361.
21. Kepler, T., and T. Elston. 2000. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* 81:3116–3136.

22. Hasty, J., J. Pradines, …, J. Collins. 2000. Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci. USA.* 97:2075–2080.

23. Gardner, T., C. Cantor, and J. Collins. 2000. Construction of a genetic toggle switch in *Escherichia coli*. *Nature.* 403:339–342.

24. Segal, E., and J. Widom. 2009. From DNA sequence to transcriptional behaviour: a quantitative approach. *Nat. Rev. Genet.* 10:443–456.

25. Sasai, M., and P. G. Wolynes. 2003. Stochastic gene expression as a many-body problem. *Proc. Natl. Acad. Sci. USA.* 100:2374–2379.

26. Sasai, M., Y. Kawabata, …, T. Terada. 2013. Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells. *PLOS Comput. Biol.* 9:e1003380.

27. Li, C., and J. Wang. 2013. Quantifying Waddington landscapes and paths of non-adiabatic cell fate decisions for differentiation, reprogramming and transdifferentiation. *J. R. Soc. Interface.* 10: 20130787.

28. Feng, H., B. Han, and J. Wang. 2012. Landscape and global stability of nonadiabatic and adiabatic oscillations in a gene network. *Biophys. J.* 102:1001–1010.

29. Ge, H., H. Qian, and X. S. Xie. 2015. Stochastic phenotype transition of a single cell in an intermediate region of gene state switching. *Phys. Rev. Lett.* 114:078101.

30. Walczak, A. M., J. N. Onuchic, and P. G. Wolynes. 2005. Absolute rate theories of epigenetic stability. *Proc. Natl. Acad. Sci. USA.* 102:18926–18931.

31. Feng, H., and J. Wang. 2012. A new mechanism of stem cell differentiation through slow binding/unbinding of regulators to genes. *Sci. Rep.* 2:550.

32. Lu, M., J. Onuchic, and E. Ben-Jacob. 2014. Construction of an effective landscape for multistate genetic switches. *Phys. Rev. Lett.* 113:078102.

33. Roma, D. M., R. A. O'Flanagan, …, R. Mukhopadhyay. 2005. Optimal path to epigenetic switching. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 71:011902.

34. Aurell, E., and K. Sneppen. 2002. Epigenetics as a first exit problem. *Phys. Rev. Lett.* 88:048101.

35. Zhou, J., M. Aliyu, …, S. Huang. 2012. Quasi-potential landscape in complex multi-stable systems. *J. R. Soc. Interface.* 9:3539–3553.

36. Wang, P., C. Song, …, J. Xing. 2014. Epigenetic state network approach for describing cell phenotypic transitions. *Interface Focus.* 4:20130068.

37. Zhang, K., M. Sasai, and J. Wang. 2013. Eddy current and coupled landscapes for nonadiabatic and nonequilibrium complex system dynamics. *Proc. Natl. Acad. Sci. USA.* 110:14930–14935.

38. Morelli, M., S. Tanase-Nicola, …, P. ten Wolde. 2008. Reaction coordinates for the flipping of genetic switches. *Biophys. J.* 94:3413–3423.

39. Strasser, M., F. J. Theis, and C. Marr. 2012. Stability and multiattractor dynamics of a toggle switch based on a two-stage model of stochastic gene expression. *Biophys. J.* 102:19–29.

40. Schultz, D., A. M. Walczak, …, P. G. Wolynes. 2008. Extinction and resurrection in gene networks. *Proc. Natl. Acad. Sci. USA.* 105: 19165–19170.

41. Lipshtat, A., A. Loinger, …, O. Biham. 2006. Genetic toggle switch without cooperative binding. *Phys. Rev. Lett.* 96:188101.

42. Ma, R., J. Wang, …, H. Liu. 2012. Small-number effects: a third stable state in a genetic bistable toggle switch. *Phys. Rev. Lett.* 109: 248107.

43. Artyomov, M. N., M. Mathur, …, A. K. Chakraborty. 2009. Stochastic bimodalities in deterministically monostable reversible chemical networks due to network topology reduction. *J. Chem. Phys.* 131:195103.

44. Allen, R., C. Valeriani, and P. ten Wolde. 2009. Forward flux sampling for rare event simulations. *J. Phys. Condens. Matter.* 21:463102.

45. Dickson, A., A. Warmflash, and A. R. Dinner. 2009. Nonequilibrium umbrella sampling in spaces of many order parameters. *J. Chem. Phys.* 130:074104.

46. Donovan, R., A. Sedgewick, …, D. Zuckerman. 2013. Efficient stochastic simulation of chemical kinetics networks using a weighted ensemble of trajectories. *J. Chem. Phys.* 139:115105.

47. Schütte, C., F. Noé, …, E. Vanden-Eijnden. 2011. Markov state models based on milestoning. *J. Chem. Phys.* 134:204105.

48. Dickson, A., A. Warmflash, and A. R. Dinner. 2009. Separating forward and backward pathways in nonequilibrium umbrella sampling. *J. Chem. Phys.* 131:154104.

49. Vanden-Eijnden, E., and M. Venturoli. 2009. Revisiting the finite temperature string method for the calculation of reaction tubes and free energies. *J. Chem. Phys.* 130:194103.

50. Adelman, J., and M. Grabe. 2013. Simulating rare events using a weighted ensemble-based string method. *J. Chem. Phys.* 138:044105.

51. Ptashne, M. 2004. A Genetic Switch, 3rd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

52. Niwa, H., Y. Toyooka, …, J. Rossant. 2005. Interaction between Oct3/4 and Cdx2 determines trophectoderm differentiation. *Cell.* 123: 917–929.

53. Zhang, P., G. Behre, …, Z. Sun. 1999. Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1. *Proc. Natl. Acad. Sci. USA.* 96:8705–8710.

54. Hong, T., J. Xing, …, J. Tyson. 2012. A simple theoretical framework for understanding heterogeneous differentiation of CD4+ T cells. *BMC Syst. Biol.* 6:66.

55. Waddington, C., and H. Kacser. 1957. The Strategy of the Genes. Routledge, London.

56. Huang, S. 2012. The molecular and mathematical basis of Waddington's epigenetic landscape: a framework for post-Darwinian biology? *BioEssays.* 34:149–157.

57. Huber, G., and S. Kim. 1996. Weighted-ensemble Brownian dynamics simulations for protein association reactions. *Biophys. J.* 70:97–110.

58. Ren, E. W. W., and E. Vanden-Eijnden. 2005. Transition pathways in complex systems: reaction coordinates, isocommittor surfaces, and transition tubes. *Chem. Phys. Lett.* 413:242–247.

59. Zwier, M. C., and L. T. Chong. 2010. Reaching biological timescales with all-atom molecular dynamics simulations. *Curr. Opin. Pharmacol.* 10:745–752.

60. Allen, R. J., P. B. Warren, and P. R. ten Wolde. 2005. Sampling rare switching events in biochemical networks. *Phys. Rev. Lett.* 94:018104.

61. Dickson, A., and A. R. Dinner. 2010. Enhanced sampling of nonequilibrium steady states. *Annu. Rev. Phys. Chem.* 61:441–459.

62. Van Kampen, N. G. 2007. Stochastic Processes in Physics and Chemistry, 3rd ed. Elsevier, Amsterdam.

63. MATLAB and Parallel Computing Toolbox Release. 2012b. The MathWorks, Natick, MA.

64. Faeder, J. R., M. L. Blinov, and W. S. Hlavacek. 2009. Rule-based modeling of biochemical systems with BioNetGen. *Methods Mol. Biol.* 500:113–167.

65. Zwier, M. C., J. L. Adelman, …, L. T. Chong. 2015. WESTPA: An interoperable, highly scalable software package for weighted ensemble simulation and analysis. *J. Chem. Theory Comput.* 11:800–809.

66. Ma, A., and A. R. Dinner. 2005. Automatic method for identifying reaction coordinates in complex systems. *J. Phys. Chem. B.* 109:6769–6779.

67. Yamanaka, Y., F. Lanner, and J. Rossant. 2010. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development.* 137:715–724.

68. Singh, A. M., T. Hamazaki, …, N. Terada. 2007. A heterogeneous expression pattern for Nanog in embryonic stem cells. *Stem Cells.* 25:2534–2542.

69. Chambers, I., J. Silva, …, A. Smith. 2007. Nanog safeguards pluripotency and mediates germline development. *Nature.* 450:1230–1234.

70. Kalmar, T., C. Lim, …, A. Martinez Arias. 2009. Regulated fluctuations in Nanog expression mediate cell fate decisions in embryonic stem cells. *PLoS Biol.* 7:e1000149.

71. Buganim, Y., D. A. Faddah, …, R. Jaenisch. 2012. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. *Cell.* 150:1209–1222.

72. Mariani, L., E. Schulz, …, T. Höfer. 2010. Short-term memory in gene induction reveals the regulatory principle behind stochastic IL-4 expression. *Mol. Syst. Biol.* 6:359.

73. Borrero, E. E., and F. A. Escobedo. 2007. Reaction coordinates and transition pathways of rare events via forward flux sampling. *J. Chem. Phys.* 127:164101.

74. Ren, E. W. W., and E. Vanden-Eijnden. 2002. String method for the study of rare events. *Phys. Rev. B.* 66:052301.