

# Predicting rates of cell state change caused by stochastic fluctuations using a data-driven landscape model

Daniel R. Sisan, Michael Halter, Joseph B. Hubbard, and Anne L. Plant<sup>1</sup>

National Institute of Standards and Technology, Gaithersburg, MD 20899

Edited\* by Peter G. Wolynes, Rice University, Houston, TX, and approved October 2, 2012 (received for review May 16, 2012)

**We develop a potential landscape approach to quantitatively describe experimental data from a fibroblast cell line that exhibits a wide range of GFP expression levels under the control of the promoter for tenascin-C. Time-lapse live-cell microscopy provides data about short-term fluctuations in promoter activity, and flow cytometry measurements provide data about the long-term kinetics, because isolated subpopulations of cells relax from a relatively narrow distribution of GFP expression back to the original broad distribution of responses. The landscape is obtained from the steady state distribution of GFP expression and connected to a potential-like function using a stochastic differential equation description (Langevin/Fokker-Planck). The range of cell states is constrained by a force that is proportional to the gradient of the potential, and biochemical noise causes movement of cells within the landscape. Analyzing the mean square displacement of GFP intensity changes in live cells indicates that these fluctuations are described by a single diffusion constant in log GFP space. This finding allows application of the Kramers' model to calculate rates of switching between two attractor states and enables an accurate simulation of the dynamics of relaxation back to the steady state with no adjustable parameters. With this approach, it is possible to use the steady state distribution of phenotypes and a quantitative description of the short-term fluctuations in individual cells to accurately predict the rates at which different phenotypes will arise from an isolated subpopulation of cells.**

population distribution | dynamical systems | stochastic protein expression | biological noise

Genetically identical cells do not respond identically when exposed to nominally identical environmental conditions. Such nongenetic phenotypic variability has been widely observed in bacteria (1, 2), yeast (3), and mammalian cells (4–8). Population heterogeneity is thought to result from the inherently stochastic nature of intracellular events, which are subject to statistical fluctuations caused by small copy numbers of the constituent molecules, such as transcription factors (9). Many investigations into the origins and effects of stochastic gene expression have used engineered organisms and stochastic gene network models to determine the sources and magnitude of variability (10–12). These fluctuations, although causing continual change at the single-cell level, can lead to stable distributions of phenotypes within a population.

The idea of a stable distribution of states in the presence of random fluctuations is reminiscent of statistical physics, where randomness results from thermal fluctuations and the stable distribution of states reflects a potential energy function. The popular concept of the epigenetic landscape suggested in the work by Waddington (13) (i.e., a surface of branching valleys and ridges on which cells explore phenotypic states) can be thought of as a series of potential energy functions. The epigenetic landscape, in which different phenotypic states arise, despite the fact that cells have identical gene sequences, has been discussed widely, mostly in the context of developmental biology and stem cell fate decisions (14–17); it has been a useful framework, even as a qualitative description. Several studies have focused on developing the landscape concept quantitatively, where the connection between the

landscape and an energy-like potential is made explicit and rigorous, but so far, these efforts have been limited to either theoretical treatments (18, 19) or completely specified systems of chemical reactions (20, 21) and have not been tested with experimental data.

Here, we develop an approach that allows us to quantitatively describe a landscape of cell phenotypes based on experimental data, and it is applicable to cell systems where the relevant underlying biochemistry is unknown or difficult to measure. The approach is similar to the landscape models of protein folding (22, 23) in several ways: the complexity of the system precludes a complete description, we focus on the dynamics of the system, and we carefully consider the choice of reaction coordinate (24). Our reaction coordinate is a 1D space (the  $x$  axis of the landscape), in which entities move diffusively and are subject to nonrandom forces determined by the gradient of the potential.

In this paper, we examine a fibroblast cell line that is stably transfected to express GFP in response to activation of the promoter for the ECM protein, tenascin-C (TN-C). TN-C, which is controlled by a large promoter sequence with a number of transcription factor binding sites (Fig. S1), is highly regulated both temporally and spatially during development, and in the adult, it is expressed predominantly under conditions of wound healing and tumor growth (25–27) and in hypertensive arteries (28), where it supports vascular smooth muscle cell proliferation, migration, and survival (29, 30). In our experiments, a clonal population of cells is grown under homogeneous conditions but exhibits a wide range of GFP intensities, most likely because of noise in promoter activity. To probe the dynamics underlying this variability, we use two types of kinetic experiments. One type is time-lapse microscopy to quantify fluctuations in GFP intensity in individual living cells. The second type isolates subpopulations of cells by cell sorting according to their GFP intensity and follows the kinetics of relaxation of these populations as they revert from their sorted distribution back to the steady state distribution. We find that the relaxation of a subpopulation back to the steady state distribution can be partially described by a simple two-state switching model, but an accurate analysis of the kinetics of relaxation requires a continuum model. We use a Langevin-type stochastic differential equation, which leads to a 1D quantitative potential landscape. The steady state population distribution of GFP is used to derive the potential. The measured fluctuations in cellular GFP, determined by time-lapse microscopy of individual living cells, are used to determine that the appropriate reaction coordinate is log GFP concentration, in which a single, constant diffusion coefficient characterizes fluctuations in GFP. This finding allows application of the classic Kramers' theory of potential barrier crossing and prediction of the rates of switching between the two states based solely on the

Author contributions: D.R.S., M.H., and A.L.P. designed research; D.R.S. and M.H. performed research; D.R.S., M.H., and J.B.H. analyzed data; and D.R.S., M.H., J.B.H., and A.L.P. wrote the paper.

The authors declare no conflict of interest.

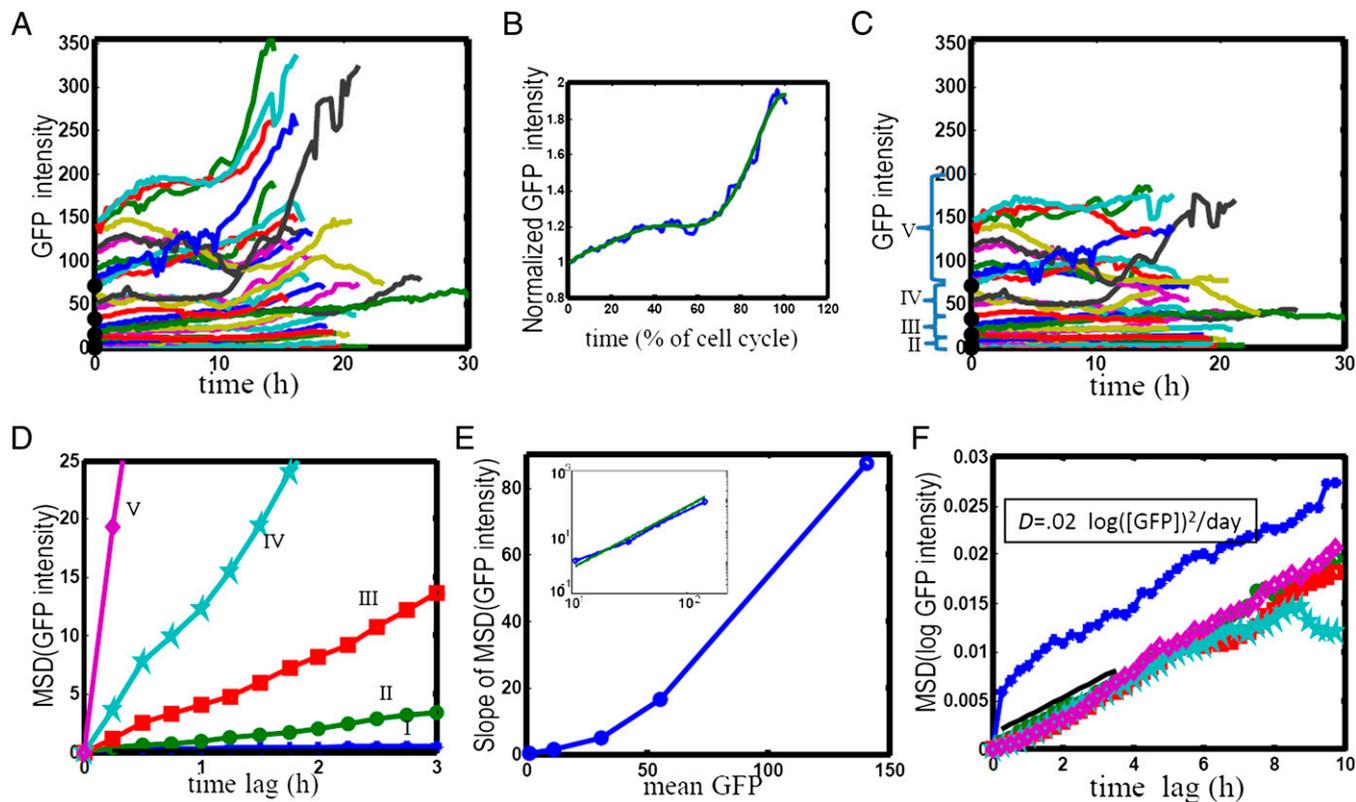
\*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence should be addressed. E-mail: anne.plant@nist.gov.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1207544109/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1207544109/-DCSupplemental).





**Fig. 2.** Trajectories of cellular GFP intensities are diffusive, with a diffusion coefficient that is constant in log GFP space. (A) GFP intensity over time is shown for 32 representative cells (of 344 total analyzed) from mitosis to mitosis obtained from automated quantitative live-cell segmentation and tracking (*SI Text, section 3*). (B) The average trajectory (synchronized as in Fig. 1 and then normalized). The green curve is a sixth-order polynomial fit used as a smooth approximation to the data. (C) The detrended GFP intensity trajectories from A after dividing each cell's trajectory by the fit in B. (D) The MSD is calculated for each detrended trajectory  $\Delta t$ , and then, it is grouped into five bins (shown in C) according to GFP intensity and averaged. Only a few data points for groups IV and V are shown in the graph. (E) The slope increases approximately as the square of GFP intensity. *Inset* shows the same plot on log-log axes superimposed with a line of slope 2, the expected form for a square dependence on GFP. (F) The slopes of the MSD are approximately the same when calculated using the log-transformed trajectories in C. The MSD for the lowest intensity bin (blue) has a higher offset from the rest because of a higher relative measurement noise, which we can estimate to be <1% of the average GFP intensity for the population.

thought of as describing a random walk in GFP concentration space, which is subject to random fluctuations that arise from small copy numbers of chemical components, like transcription factors, and an average nonrandom force,  $A(x)$ , which describes the deterministic rules arising from the signaling pathways that control GFP expression levels.

Eq. 1 is not well-defined when the diffusion coefficient,  $D(x)$ , depends on the protein concentration,  $x$ . More than a single Fokker-Planck equation can be derived depending on the interpretation of the stochastic process (*SI Text, section 4*). In cases where  $D(x)$  depends on the reaction coordinate, a Langevin equation with a constant diffusion coefficient can often be derived by transforming the reaction coordinate of Eq. 1 to the following (unique) reaction coordinate (32) (Eq. 2):

$$\bar{x} = \int_0^x \frac{dx}{\sqrt{D(x)}}, \quad [2]$$

where  $\bar{x}$  is a function of  $x$ , and as a result, the diffusion coefficient,  $\bar{D}$ , becomes a new variable that is independent of  $x$ . This operation transforms Eq. 1 into the following Langevin equation (Eq. 3):

$$\frac{d\bar{x}}{dt} = \bar{A}(\bar{x}) + \sqrt{2\bar{D}}\eta(t). \quad [3]$$

With this transformation, the force,  $\bar{A}(\bar{x})$ , now depends on the functional forms of  $D(x)$  and  $A(x)$ —and on the choice of

interpretation. Therefore, the dilemma of needing to make an interpretation has not been avoided but has shifted from the noise term to  $\bar{A}(\bar{x})$ .

We now show that the dilemma can be totally eliminated, because we can determine  $\bar{A}(\bar{x})$  empirically from the measured steady state distribution of protein concentration as follows. With a constant diffusion coefficient, the Langevin equation can be used to construct the following Fokker-Planck equation for the probability distribution  $P(\bar{x}, t)$  (33) describing a population of cells (Eq. 4):

$$\partial_t P(\bar{x}, t) = -\partial_{\bar{x}} \left( \bar{A}(\bar{x}) P(\bar{x}, t) \right) + \frac{\bar{D}}{2} \partial_{\bar{x}}^2 P(\bar{x}, t). \quad [4]$$

The steady state probability distribution,  $P_{ss}$ , can be solved by setting the left-hand side to zero and integrating, and it is given by (Eq. 5)

$$P_{ss}(\bar{x}) = C \exp \left( \frac{2}{\bar{D}} \int_0^{\bar{x}} \bar{A}(\bar{x}') d\bar{x}' \right), \quad [5]$$

where  $C$  is a normalization constant. We now define the following function (Eq. 6),

$$U(\bar{x}) = -\int_0^{\bar{x}} \bar{A}(y) dy, \quad [6]$$

which we call a potential, because it puts the steady state distribution  $P_{ss}(\bar{x}) \sim \exp(-2U(\bar{x})/\bar{D})$  in a form that is analogous to the Boltzmann distribution  $P \sim \exp(-U/k_B T)$ . Furthermore, the force  $\bar{A}(\bar{x}) = -\frac{\partial U}{\partial \bar{x}}$  is proportional to the gradient of the potential. Importantly, the potential and the force can be obtained from the experimentally accessible steady state probability distribution,  $P_{ss}(\bar{x})$ , and the diffusion coefficient,  $\bar{D}$  (Eq. 7):

$$U(\bar{x}) = -\left(\frac{\bar{D}}{2}\right) \ln(P_{ss}(\bar{x})) \quad [7]$$

and (Eq. 8)

$$\bar{A}(\bar{x}) = -\frac{\partial U}{\partial \bar{x}} = \frac{\bar{D}}{2} \frac{\partial \ln(P_{ss}(\bar{x}))}{\partial \bar{x}}. \quad [8]$$

This allows construction of a self-consistent, unambiguous Langevin description (and simulation) without direct knowledge of  $A(x)$  and without choosing between an Ito or Stratonovich interpretation (although which interpretation is appropriate remains an open question). In approaches that start with  $A(x)$  instead of empirical data (20), a choice between interpretations must be made and justified, because each interpretation leads to a different outcome.

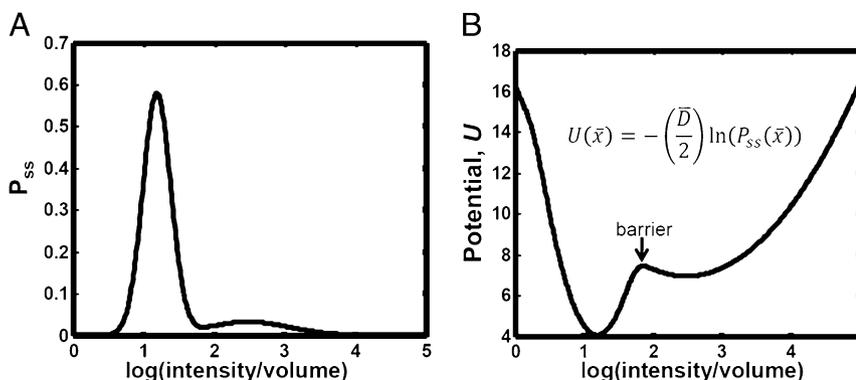
**Obtaining the Diffusion Coefficient and Potential Landscape from Cellular Data.** A complete mathematical description of the steady state distribution is required for establishing the potential landscape, and this description requires determining the appropriate reaction coordinate. An analysis of the noise in cellular GFP intensity is used to obtain  $\bar{x}$ , the appropriate reaction coordinate for the landscape, and  $D(x)$ , the diffusion coefficient, with which cells explore the landscape. This analysis is shown in Fig. 2. Representative fluorescence trajectories of total intensity in individual cells from time-lapse live-cell microscopy are shown in Fig. 2A. To approximate the concentration and minimize the cell cycle-dependent trends, each trajectory was detrended using the fit to the average trend in Fig. 2B, which is shown in Fig. 2C. The mean squared displacement (MSD) was calculated as  $MSD(\Delta t) = \langle [x(t + \Delta t) - x(t)]^2 \rangle_t$ , where  $x$  is detrended GFP intensity,  $\Delta t$  is the time lag, and the bracket denotes an average over  $\Delta t$ . An MSD is determined at all  $\Delta t$  values for each cell, and then, MSD values are binned according to average cell intensity as indicated in Fig. 2C. As seen in Fig. 2D, all curves were found to be linear at short times, indicating that the dynamics of fluctuation in GFP are diffusive and therefore, that the Langevin equation is an appropriate description. The slope of the MSD, from which we derive the diffusion coefficient  $D = slope/2$ , depends strongly on the average GFP intensity of the cells (Fig. 2E), indicating GFP-dependent variations in the diffusion coefficient. In particular, the slope and thus, the diffusion coefficient increase as the square of the mean GFP.

We can use this relationship to determine a reaction coordinate in which the diffusion coefficient is constant. Substituting  $D(x) = x^2$  into Eq. 2,  $\bar{x} = \int_0^{\bar{x}} \frac{dx}{x} = \log x$ , indicates that the log-transformed GFP concentration should have fluctuations characterized by a constant diffusion coefficient, which is verified in Fig. 2F. The slopes of the MSD are approximately the same, indicating a constant diffusion coefficient,  $D = 0.02 \pm 0.002 \log^2(\text{GFP})/d$  (a description of error calculations is given in *SI Text, section 3*). Thus, we choose log GFP concentration as the reaction coordinate.

The steady state distribution,  $P_{ss}(\bar{x})$ , can be transformed by Eq. 6 to result in the potential landscape. The steady state distribution shown in Fig. 3A is derived from the observed steady state distribution (Fig. S24) and has been modified using the switching rates determined from the two-state model (Fig. S3) to deduce the steady state distribution resulting solely from variability of promoter activity decoupled from the effects of population dynamics (*SI Text, section 1*). The landscape,  $U(\bar{x})$ , which is derived from the steady state distribution, is shown in Fig. 3B. The landscape consists of two minima, which correspond to the two major states (attractors). The stability of the states is determined by the depth of each minimum and the height of the barrier. The diffusion coefficient,  $D$ , characterizes how fast individual cells in the population explore the range of phenotypes within each state.

**Langevin Simulation.** Stochastic numerical simulations of an ensemble of Langevin equations were performed (*SI Text, section 2* and Fig. S4) to simulate the relaxation of the four sorted dishes (Fig. 1 and Fig. S3), with the force given by  $\bar{A}$ , the gradient of the landscape (Eq. 7), and the diffusion coefficient determined from the data in Fig. 2. In Fig. 4, the results of the simulations are compared with the experimental data from Fig. 1 and found to agree well without any tuned parameters. An analysis of the sensitivity of the simulation to different values for  $D$  is shown in Fig. S4 B–E, and it indicates that the simulation prediction is rather insensitive to changes in  $D$  up to a factor of two. This result suggests that the MSD analysis of the dynamic data, although orthogonal to the distribution measurements by flow cytometry, provides for a  $D$  that uniquely predicts the relaxation data. The simulation captures two features that were missed by the two-state description. First, for Dish II, in which the subpopulation of cells had a phenotype that was near the barrier of the two major states, the curve first increases and then decreases. Second, Dish IV takes longer to relax than Dish III, despite both being initially composed of all high-producing cells because of the time needed for cells to diffuse close to the barrier.

The different rates of proliferation for low- and high-activity cells must be included in the simulation to provide an accurate description for the same reasons as for the two-state model (Fig. S3A). Although consideration of state-dependent rates of proliferation precludes a purely analytical time-dependent solution for the probability distribution from the Fokker–Planck equation (Eq. 3), we show that it is easily incorporated into a simulation.



**Fig. 3.** The epigenetic landscape is derived from the measured steady state distribution. (A) The estimated unbiased steady state distribution was obtained by modifying the fit of the distribution in Fig. S1 to eliminate the effect of state-dependent proliferation rates by applying the ratio of low- to high-activity cells,  $\frac{k_{HL}}{k_{LH}} \approx \frac{1}{6}$ , as determined from the two-state model (*SI Text, section 1*). (B) The potential is defined by Eq. 7 [ $U(\bar{x}) = -\bar{D}/2 \ln(P_{ss}(\bar{x}))$ ], where  $P_{ss}(\bar{x})$  is the probability distribution shown in A,  $\bar{x} = \log([\text{GFP}])$ , and  $\bar{D} = 0.02 \log^2 [\text{GFP}]/d$ .

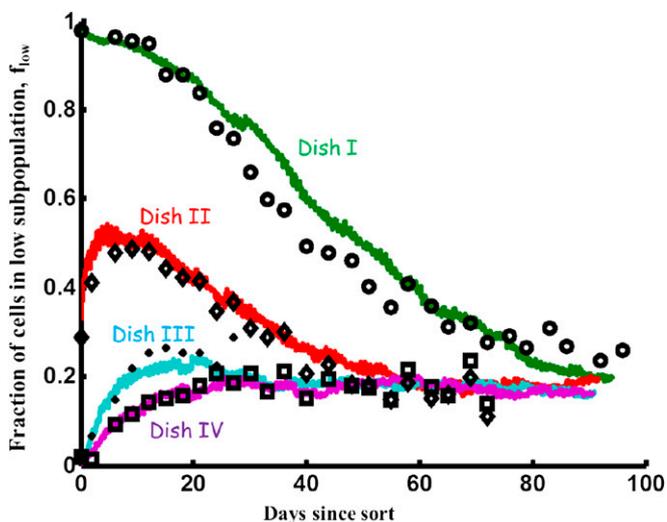
**Predicting Switching Rates Using the Potential Landscape: Kramers' Theory.** The shape of landscape shown in Fig. 3B, with its potential minima and barrier between them, allows us to directly compute the rate constants for switching between two stable states (attractors) using Kramers' classic theory of energy barrier crossing (34). Relying solely on the shape of the landscape, these rates are given by (Eq. 9)

$$K = \frac{1}{2\pi} \sqrt{U''(x_{\min})|U''(x_{\max})|} \exp(-[U(x_{\max}) - U(x_{\min})]/D), \quad [9]$$

where  $U$  is the potential function, primes denote derivatives [therefore,  $U''(x_{\min})$  is the landscape's curvature at a local minimum],  $x_{\max}$  is the position of the barrier peak, and  $D$  is the diffusion coefficient. For the Kramers' model,  $D$  must be constant. As we show above, the only transformation of the reaction coordinate that produces a constant  $D$ , in our case, the log transformation, will yield a landscape with the correct curvature for Eq. 8. Furthermore, note that the curvature implicitly includes  $D$ . Applying this formula to the potential in Fig. 3B and using  $D = 0.02 \log_2(\text{GFP})/\text{d}$ , which was obtained from Fig. 2F, we obtain  $K_{LH} = 0.004/\text{d}$  as the rate constant for transition of low-activity cells to high-activity cells and  $K_{HL} = 0.025/\text{d}$  for the transition of high-activity cells to low-activity cells. These values agree to within 50% of the values estimated from the fitting of the relaxation data to the two-state model (Fig. S3), and establish the predictive use of the landscape.

## Discussion

We have shown how to rigorously derive the potential landscape for the activity of a gene promoter. The landscape is derived from the steady state distribution of the range of promoter activities within the population as determined by flow cytometry or static imaging measurements, and the kinetics of fluctuations in promoter activity in individual cells are determined from time-dependent fluorescence imaging data. Our approach is based on a Langevin equation, which includes the force that defines the shape of the landscape and a single kinetic constant that describes the fluctuations in promoter activity. The model accurately predicts the complex kinetics with which cells that are sorted according to their promoter activity relax back to the steady state distribution of



**Fig. 4.** Results of a Langevin simulation of the relaxation experiment. Data points are the same as in Fig. S2. The simulation uses the experimentally derived diffusion coefficient,  $\bar{D}$ , the landscape,  $U(\bar{x})$  (Fig. 3B), and state-dependent rates of proliferation. Each time point in the simulation produces a population distribution of GFP concentrations, which is reduced to a scalar: the fraction of cells below a GFP concentration threshold as in Fig. S2 and described in *S1 Text*, section 3.

activities. The time required for a selected subpopulation to transition from its narrow distribution of states back to a broad distribution of phenotypes can take weeks, but it can be predicted using data that can be acquired in about 1 d. A similar slow relaxation was observed for subpopulations of stem cells sorted for Sca-1 activity, and the work by Chang et al. (8) showed the significance of population heterogeneity to differentiation potential. In that report, the data were modeled as a transition between two states (8). In this report, we also explored a two-state model, but in addition, we performed a kinetic analysis of fluctuations in promoter activity using time-lapse microscopy on individual cells. These data established that these fluctuations are diffusive in nature. The transformation of the reaction coordinate to log GFP provides a constant diffusion coefficient that allowed us to use a continuum model in the form of a simple Langevin equation and the Kramers' theory to predict rates of transition between states.

The TN-C promoter seems to have two minima in its potential landscape. In a microarray gene expression study using hundreds of tissue samples in mice (35), TN-C was reported to be one of the genes that is expressed with a bimodal distribution of levels. It was reported to be off more often than on, consistent with our deduced steady state distribution, suggesting that the activity of the promoter measured in cell culture reflects, at least partially, its physiological activity. The potential landscape that we determine allows the switching rates between these two relatively stable promoter activity states to be predicted by applying Kramers' theory (34), which takes into account the shape of the landscape and the diffusion coefficient. The calculated Kramers' switching rates,  $K_{LH} = 0.004/\text{d}$  and  $K_{HL} = 0.025/\text{d}$ , were found to agree to within 50% of the switching rates obtained from the sorting/relaxation experiment,  $k_{LH} = 0.004/\text{d}$  and  $k_{HL} = 0.017/\text{d}$ . Based on these results, we can predict that, at the single-cell level (ignoring progeny from division), one would wait, on average,  $k_{HL}^{-1} \approx 50$  d for a high-activity cell to switch and  $k_{LH}^{-1} \approx 300$  d for a low-activity cell to switch. Considering progeny and starting with a single cell that doubles approximately every 1 d, one would wait, on average,  $\log_2(50) \approx 6$  d (for a high cell) and  $\log_2(300) \approx 8$  d (for a low cell) before observing a cell that had switched. The predictive value of this approach to modeling cellular dynamics may have important practical implications for applications in stem cell biology and biomanufacturing as well as tumor heterogeneity and drug resistance.

Although the Kramers' rate of transition of cells in the population from the low to high states agrees with the values found from the two-state model, it should be noted that the two-state model only provides a reasonable fit to the relaxation responses from the populations of cells with the lowest and highest promoter activities. The intermediate populations show more complex relaxation kinetics, which require the Langevin/Fokker-Planck model to describe adequately. The biological significance of these findings is that switching from a low- to high-activity state for TN-C promoter activity is a diffusive process; therefore, cells with concentrations near the state barrier are closer to switching than cells away from the barrier. Cells that are near the barrier have a finite probability of changing in activity in either direction, thus displaying a more complex response in activity with time of relaxation. An alternate outcome could have been that switching occurs spontaneously through some discrete process without coupling to a slow variation because of noise in response within the population. However, our results (e.g., the faster relaxation of Dish III vs. Dish IV in Fig. 4) show that this result is not the case for TN-C, and the simulation allows us to predict how cells explore state space differently depending on where they reside in the landscape. A practical implication of this finding is that, for cell-sorting experiments, the details of the window size and location used to select subpopulations can have a measurable influence on the apparent stability of the selected subpopulations.

Fluctuations in promoter activity, caused by binding and unbinding of effector molecules, can dominate protein number fluctuations (36). For TN-C, the dependence of the noise in GFP on the intensity of GFP expression suggests a multiplicative process in the activation of the promoter. This finding is consistent with the known characteristics of the TN-C promoter, which is

representative of a class of environmentally responsive proteins (*SI Text, section 6*).

It is important to note that, in a population of cells that exhibits a stationary distribution of phenotypic states, no state is a permanent condition for any cell. A force constrains the state space that can be explored by cells; however, because the entire space is never explored by any individual cell during its lifetime, the Langevin dynamics of the population by necessity require mother/daughter correlations (which we have independently determined). Those multigenerational correlations are epigenetic in nature in the broadest sense; they could be caused by epigenetic modifications to the DNA or perhaps, passage of high concentrations of transcription factors from mothers to daughters.

In this work, we have correctly predicted the time dependence for relaxation of subpopulations of cells as they reestablish the steady state distribution of promoter activities. This approach will likely be applicable to other similar perturbation/relaxation scenarios involving other cellular phenotypic features. For example, the induction of differentiation of a pluripotent population of cells may be thought of as imposing, through chemical treatment, a new potential landscape of gene expression on that population. The rate at which the pluripotent population relaxes to this new potential landscape should be predictable from the landscape that is derived from the final steady state distribution and the rate of fluctuation of the phenotypic feature.

## Materials and Methods

**Cells, Cell Culture, Quantitative Imaging, and Image Analysis.** The preparation of the cell line and all cell culture, imaging, and image analysis were performed as previously published (31). NIH 3T3 mouse fibroblasts (ATCC) were transfected with a construct of a 4.1-kbp fragment of the TN-C promoter (provided by Peter L. Jones, University of Pennsylvania, State College, PA) fused to a destabilized EGFP (Clontech Laboratories). Phase contrast and fluorescence images were acquired every 15 min for >62 h in 36 different fields. Additional details can be found in *SI Text, section 5*.

**Flow Cytometry and FACS.** Flow cytometry was performed with a Beckman Coulter Quanta SC Flow Cytometer. Volume and intensity measurements were performed simultaneously for each cell to obtain the relative GFP concentration (intensity/volume). FACS was performed using a FACSARIA II Sorter (BD Biosciences). Cells were first gated on 2D forward and side scatter to exclude debris followed by four gates based on the FITC (GFP) channel fluorescence to isolate four subpopulations. Additional details can be found in *SI Text, section 5*.

**Steady State Distribution Fitting.** The steady state distribution of GFP concentration for cells in the population (Fig. S2A) as measured by flow cytometry was fit to a sum of two log normals by log-transforming the data, binning, and then fitting the histogram to a sum of two Gaussians:  $f(x) = a_1 \frac{1}{\sqrt{2\pi\sigma_1^2}} \exp[-(x - \mu_1)^2 / 2\sigma_1^2] + a_2 \frac{1}{\sqrt{2\pi\sigma_2^2}} \exp[-(x - \mu_2)^2 / 2\sigma_2^2]$ , where  $x = \log(\text{intensity/volume})$ , and  $a_1, a_2, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2$  are free parameters. The fitting was performed using the nonlinear curve-fitting function `lsqcurvefit` in Matlab (Optimization Toolbox). The parameters  $a_1, a_2$  give the relative weighting of each Gaussian. The best-fit values in Fig. S2A were  $a_1 = 0.33, a_2 = 0.67, \mu_1 = 4.2, \mu_2 = 5.5, \sigma_1^2 = 0.19,$  and  $\sigma_2^2 = 0.62$ . For Fig. 3A, the weightings were changed to be  $a_1 = 0.86$  and  $a_2 = 0.14$ ; therefore,  $\frac{a_2}{a_1} = \frac{k_{LU}}{k_{UH}} = \frac{1}{6}$  and the parameters used were  $\mu_1 = 1.2, \mu_2 = 2.5, \sigma_1 = 0.20,$  and  $\sigma_2 = 0.58$ .

**Quantifying the Portion of Cells in Low State.** For Fig. 4 and Fig. S3, the histograms measured by flow cytometry are reduced to a single scalar: the fraction of cells in the low state, which we determine by a threshold. The threshold was chosen to be the local minimum (between the two peaks) of the log-transformed GFP concentration. The portion of cells above and below this threshold was found to well approximate  $a_1$  and  $a_2$  from fitting (Fig. S2A).

**ACKNOWLEDGMENTS.** This work was funded, in part, by a National Institute of Standards and Technology National Research Council (NRC) Postdoctoral Fellowship (to D.R.S.). Any mention of commercial products is for information only; it does not imply recommendation or endorsement by NIST.

- Novick A, Weiner M (1957) Enzyme induction as an all-or-none phenomenon. *Proc Natl Acad Sci USA* 43(7):553–566.
- Spudich JL, Koshland DE, Jr. (1976) Non-genetic individuality: Chance in the single cell. *Nature* 262(5568):467–471.
- Newman JR, et al. (2006) Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441(7095):840–846.
- Sigal A, et al. (2006) Variability and memory of protein levels in human cells. *Nature* 444(7119):643–646.
- Langenbach KJ, Elliott JT, Tona A, McDaniel D, Plant AL (2006) Thin films of Type 1 collagen for cell by cell analysis of morphology and tenascin-C promoter activity. *BMC Biotechnol* 6:14.
- Spencer SL, Gaudet S, Albeck JG, Burke JM, Sorger PK (2009) Non-genetic origins of cell-to-cell variability in TRAIL-induced apoptosis. *Nature* 459(7245):428–432.
- Stockholm D, et al. (2010) Bistable cell fate specification as a result of stochastic fluctuations and collective spatial cell behaviour. *PLoS One* 5(12):e14441.
- Chang HH, Hemberg M, Barahona M, Ingber DE, Huang S (2008) Transcriptome-wide noise controls lineage choice in mammalian progenitor cells. *Nature* 453(7194):544–547.
- Raj A, van Oudenaarden A (2008) Nature, nurture, or chance: Stochastic gene expression and its consequences. *Cell* 135(2):216–226.
- Blake WJ, KAERN M, Cantor CR, Collins JJ (2003) Noise in eukaryotic gene expression. *Nature* 422(6932):633–637.
- Mettetal JT, Muzzey D, Pedraza JM, Ozbudak EM, van Oudenaarden A (2006) Predicting stochastic gene expression dynamics in single cells. *Proc Natl Acad Sci USA* 103(19):7304–7309.
- Thattai M, van Oudenaarden A (2001) Intrinsic noise in gene regulatory networks. *Proc Natl Acad Sci USA* 98(15):8614–8619.
- Waddington HK (1957) *The Strategy of the Genes; A Discussion of Some Aspects of Theoretical Biology* (Macmillan, New York).
- Goldberg AD, Allis CD, Bernstein E (2007) Epigenetics: A landscape takes shape. *Cell* 128(4):635–638.
- MacArthur BD, Please CP, Oreffo RO (2008) Stochasticity and the molecular mechanisms of induced pluripotency. *PLoS One* 3(8):e3086.
- Slack JM (2002) Conrad Hal Waddington: The last Renaissance biologist? *Nat Rev Genet* 3(11):889–895.
- Gilbert SF (1991) Epigenetic landscaping: Waddington's use of cell fate bifurcation diagrams. *Biol Philos* 6(2):135–154.
- Wang J, Zhang K, Xu L, Wang E (2011) Quantifying the Waddington landscape and biological paths for development and differentiation. *Proc Natl Acad Sci USA* 108(20):8257–8262.
- Sasai M, Wolynes PG (2003) Stochastic gene expression as a many-body problem. *Proc Natl Acad Sci USA* 100(5):2374–2379.
- Hasty J, Pradines J, Dolnik M, Collins JJ (2000) Noise-based switches and amplifiers for gene expression. *Proc Natl Acad Sci USA* 97(5):2075–2080.
- Kim KY, Wang J (2007) Potential energy landscape and robustness of a gene regulatory network: Toggle switch. *PLoS Comput Biol* 3(3):e60.
- Bryngelson JD, Onuchic JN, Soccì ND, Wolynes PG (1995) Funnels, pathways, and the energy landscape of protein folding: A synthesis. *Proteins* 21(3):167–195.
- Onuchic JN, Luthey-Schulten Z, Wolynes PG (1997) Theory of protein folding: The energy landscape perspective. *Annu Rev Phys Chem* 48:545–600.
- Best RB, Hummer G (2010) Coordinate-dependent diffusion in protein folding. *Proc Natl Acad Sci USA* 107(3):1088–1093.
- Erickson HP, Bourdon MA (1989) Tenascin: An extracellular matrix protein prominent in specialized embryonic tissues and tumors. *Annu Rev Cell Biol* 5:71–92.
- Chiquet-Ehrismann R, Hagios C, Schenk S (1995) The complexity in regulating the expression of tenascins. *Bioessays* 17(10):873–878.
- Faissner A (1997) The tenascin gene family in axon growth and guidance. *Cell Tissue Res* 290(2):331–341.
- Jones PL, Cowan KN, Rabinovitch M (1997) Tenascin-C, proliferation and sub-endothelial fibronectin in progressive pulmonary vascular disease. *Am J Pathol* 150(4):1349–1360.
- Jones FS, Jones PL (2000) The tenascin family of ECM glycoproteins: Structure, function, and regulation during embryonic development and tissue remodeling. *Dev Dyn* 218(2):235–259.
- Cowan KN, Jones PL, Rabinovitch M (2000) Elastase and matrix metalloproteinase inhibitors induce regression, and tenascin-C antisense prevents progression, of vascular disease. *J Clin Invest* 105(1):21–34.
- Halter M, et al. (2011) Cell cycle dependent TN-C promoter activity determined by live cell imaging. *Cytometry A* 79(3):192–202.
- Risken H (1989) *The Fokker-Planck Equation: Methods of Solution and Applications*, Springer Series in Synergetics (Springer, Berlin), 2nd Ed.
- van Kampen NG (2007) *Stochastic Processes in Physics and Chemistry* (Elsevier, Amsterdam), 3rd Ed.
- Meĭnikov VI (1991) The Kramers problem: Fifty years of development. *Phys Rep* 209(1–2):1–71.
- Ertel A, Tozeren A (2008) Switch-like genes populate cell communication pathways and are enriched for extracellular proteins. *BMC Genomics* 9:3.
- Walczak AM, Onuchic JN, Wolynes PG (2005) Absolute rate theories of epigenetic stability. *Proc Natl Acad Sci USA* 102(52):18926–18931.