doi:10.1016/j.jmb.2009.11.035 J. Mol. Biol. (2010) 396, 230–244

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

Stochastic Switching in Gene Networks Can Occur by a Single-Molecule Event or Many Molecular Steps

Paul J. Choi, X. Sunney Xie and Eugene I. Shakhnovich*

Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

Received 29 May 2009; received in revised form 9 November 2009; accepted 13 November 2009
Available online 18 November 2009

Due to regulatory feedback, biological networks can exist stably in multiple states, leading to heterogeneous phenotypes among genetically identical cells. Random fluctuations in protein numbers, tuned by specific molecular mechanisms, have been hypothesized to drive transitions between these different states. We develop a minimal theoretical framework to analyze the limits of switching in terms of simple experimental parameters. Our model identifies and distinguishes between two distinct molecular mechanisms for generating stochastic switches. In one class of switches, the stochasticity of a single-molecule event, a specific and rare molecular reaction, directly controls the macroscopic change in a cell's state. In the second class, no individual molecular event is significant, and stochasticity arises from the propagation of biochemical noise through many molecular pathways and steps. As an example, we explore switches based on protein–DNA binding fluctuations and predict relations between transcription factor kinetics, absolute switching rate, robustness, and efficiency that differentiate between switching by single-molecule events or many molecular steps. Finally, we apply our methods to recent experimental data on switching in *Escherichia coli* lactose metabolism, providing quantitative interpretations of a single-molecule switching mechanism.

Keywords: gene networks; expression noise; single molecule; stochastic switching; transcription factors

Edited by M. Levitt

Introduction

The set of possible macromolecular interactions in a cell comprises a biological network, as mapped out by protein–protein or protein–DNA networks.[1–3] The state of a network may be defined as the number and binding mode of each type of macromolecule at any given time. Although, genetically, identical cells share the same underlying network structure, the precise state of each cell's network can vary, leading to disparate phenotypes. Examples of such variation range from bistable phenotypes in prokaryotes[4–6] to extensive phenotypic differentiation occurring in multicellular organisms.[7,8] Models using ordinary differential equations[9] or Boolean networks[10] have shown that multiple stable states can arise from the dynamical properties of even simple networks.

Although several experimental works have addressed the dynamics of switching by bistable genetic networks,[11–13] for most systems, the underlying assumption that biochemical noise drives stochastic switching[14–16] has not been proven. In fact, many models of stochastic switching fail to adequately describe the intricacies of noise, such as extrinsic contributions[17,18] or non-Poissonian processes.[19,20] Thus, the precise molecular sources of stochasticity in such biological switches remain unknown. We therefore develop methods to identify and test the molecular origins or properties of stochasticity in these biological networks. In doing so, we find two classes of stochastic switching—one driven by single-molecule events and another driven by many molecular steps. This result agrees with two contrasting mechanisms suggested by recent experimental reports.[12,13]

To describe these two different mechanisms, we calculate the probability that a single occurrence of a molecular reaction, such as the dissociation of a transcription factor from a promoter, will lead to a change in the network state. In this example, one class of stochastic mechanisms requires a large number of transcription factor binding fluctuations before a switching event occurs, and any particular

*Corresponding author. E-mail address: eugene@belok.harvard.edu.

transcription factor event is relatively insignificant. In this case, stochasticity of the entire biochemical reaction network generates switching. Another class of stochastic mechanisms requires only a small number of transcription factor binding fluctuations, which means that the stochasticity of a single-molecule event, that particular binding fluctuation, directly controls the switching of the network state.

By analyzing models in this context, we can relate these two classes of stochastic switches to observations from previous studies of noise-driven switching. For example, previous studies have found that accurate models of certain biological switches require the explicit consideration of protein–DNA binding kinetics rather than the use of equilibrium binding properties alone.[21,22] Other studies have also found that switching rates calculated in models are not robust but exhibit an exponential sensitivity to input parameters,[23,24] in contrast to experimentally robust switching rates.[25] Our model shows that slow transcription factor kinetics and robustness are features of switching by single-molecule events and not by many molecular steps. By using our approach, we are also able to test other mechanistic hypotheses based on experimental data.

It is clear that detailed mechanistic models are essential not only for conceptual purposes but also for quantitatively accurate predictions. However, highly parameterized simulations or mathematically involved calculations often obscure the relationship between biological observables and switching rates. Thus, in developing our model of stochastic switching, we focused on results that are compact and interpretable in terms of experimental biological parameters, while remaining quantitatively accurate over a large range of parameters. Despite its simplicity, our minimal model naturally accommodates mechanistic features such as protein bursting, transcription factor kinetics, and extrinsic noise in a way that predicts when they should have significant roles. We find that switching by a single-molecule event may be amenable to such minimal models, while even more sophisticated models are predicted to have difficulty accurately describing switching by many molecular steps.

Results

Threshold crossing by bursting protein production

We consider the simplest bistable system consisting of a hypothetical autoregulatory gene incorporating positive feedback and develop a minimal model to analyze this situation. Our approach is to describe the dynamics of protein production in the low state of the bistable system along with a critical protein number at which positive feedback on expression level occurs. We calculate the rate at which protein number fluctuations in the low state reach this critical threshold number to determine the transition rate from the low to the high state. The assumptions we use to describe these processes are likely to be most applicable to single-cell prokaryotes, although extensions to other systems may be possible.

We start by describing the production of proteins using a bursting model, which can approximate a number of different expression dynamics, but we will specify the details of the bursting mechanism in later sections. We assume that these bursts of protein production are random and uncorrelated, occurring with a frequency of a per cell cycle. We also assume that each burst produces an instantaneous, exponentially distributed number of proteins with an average size of b proteins per burst. Thus, these two parameters, a and b, define the kinetics of protein number fluctuations in this model.

In order to construct a bistable system, we introduce a feedback mechanism that affects the frequency of bursts. When the number of proteins per cell, X, is below a threshold value of c, the burst frequency is a. Above the threshold, when $X \geq c$, we encode positive feedback by increasing the burst frequency to a^*, with $a < a^*$. The mean value of cells in the low state is ab, and for the high state, it is a^*b. Thus, for well-separated states, such that $ab \ll c \ll a^*b$ and $a^* > 1$, X will have a bimodal stationary distribution. This condition is necessary for two distinguishable, bistable phenotypes to exist. For such a system, the number of proteins for an individual cell will remain near $X = ab$ or $X = a^*b$, with a low probability of transitions between the two stable states within the time of one cell cycle. Then, the resulting bimodal distribution can be approximated as the weighted sum of two separate cell populations with distinct burst frequencies. Furthermore, if proteins are assumed to be stable,[26] so that dilution by cell division sets the effective protein degradation rate, bursting protein production for a homogeneous population of cells will generate a gamma distribution for the stationary probability distribution of X, the number of proteins per cell, with shape parameter a and scale parameter b, as shown previously.[27]

Starting with all cells in a population in the low-expression state, we calculate the transition rate to the high-expression state. In brief, for a cell with a specific number of proteins, X, we calculate the probability of any burst being large enough to cross the threshold, given a burst size b, and then calculate the probability that there is at least one successful threshold crossing per cell cycle, given a burst frequency a. We approximate the distribution in the low state to be unaltered by rare transitions to the high state. This allows us to average over the resulting probability of switching per cell cycle for cells with different initial protein numbers, weighted by the quasi-steady-state distribution in the low state. In this way (see Methods for more details), we find that the transition rate to the high-expression state is approximated by the following expression:

$$k = (c/b)^a \frac{e^{-c/b}}{\Gamma(a)} \qquad (1)$$

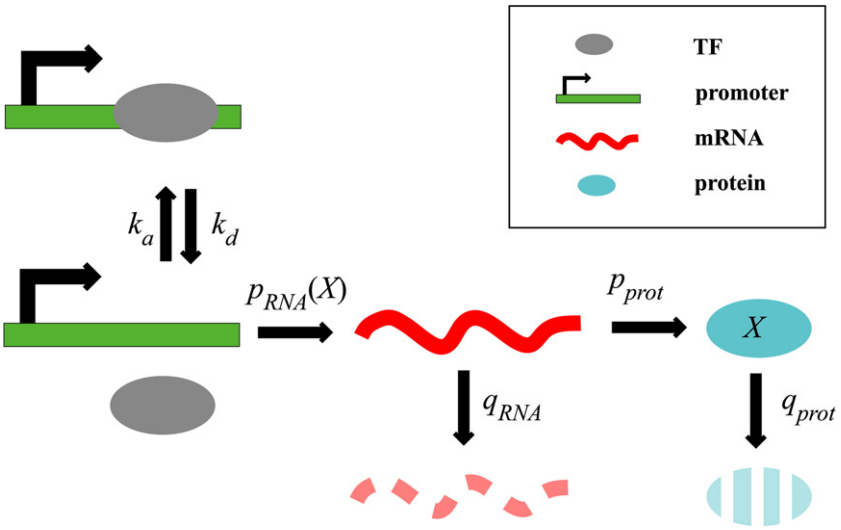

Fig. 1. Schematic of gene expression model. This model of gene expression incorporates transcription, translation, and regulation by a transcription factor (TF). Transcription only occurs from the free promoter once the repressor TF has dissociated.

We find the switching rate to be dominated by an exponential dependence for the ratio of the threshold, c, and the burst size, b, analogous to the ratio of activation energy and temperature in the Boltzmann term of chemical reaction rates. The prefactor does have a dependence on this ratio as well, though. For small values of a, the prefactor simplifies to yield $k \approx a e^{-c/b}$.

This basic model of threshold crossing by burst-like protein production allows us to explore different mechanistic aspects of stochastic switching. We will develop the interpretation of bursting in terms of transcription, translation, and transcription factor kinetics. We show that for at least some molecular mechanisms, the simplification of the underlying dynamics into three effective parameters, a, b, and c, allows us to make qualitative and quantitative predictions of switching behavior. For example, this reduction to key parameters facilitates the comparison of robustness of switching rates to perturbations for different mechanisms.

Switching by transcriptional and translational noise

The interpretation of the parameters a and b in Eq. (1) for the case of a minimal model of Poissonian transcription and translation has been described before.[27] The parameter a is the average number of mRNA transcribed per cell cycle. The parameter b is the average number of proteins translated from one mRNA. For Eq. (1) to hold in this interpretation, the mRNA lifetime should be much shorter than the protein lifetime so that the number of proteins in a cell does not depend on the current number of mRNA. Let p_{RNA} and q_{RNA} be the rates of mRNA production and degradation, respectively, and p_{prot} and q_{prot} be the corresponding rates of protein production and degradation, respectively (Fig. 1 with $k_a = 0$). If the protein is stable so that protein degradation is equated with effective dilution by cell division, q_{prot} is the rate of cell division. We will let $q_{\mathrm{prot}} = 1$ for the remainder of the discussion, which sets the time scale of the rate constants. Then, the parameters a and b have the values

$$a = p_{\mathrm{RNA}}, \quad b = p_{\mathrm{prot}} / q_{\mathrm{RNA}} \tag{2}$$

Each transcription event produces a burst of protein expression, with a size determined by the translational efficiency and mRNA lifetime. Figure 2a shows the switching rate calculated with Eqs. (1) and (2) using a fixed value of c and a range of values for a and b. We observe an asymmetric dependence of the switching rate on transcription and translation, which is more apparent in the cross sections plotted in Fig. 2b. We observe that the switching rate has a sharp dependence on the size of translational bursting, b. The switching rate has a less severe dependence on the transcription rate. Thus, the switching rate is predicted to have the greater sensitivity to factors such as mRNA lifetime or codon usage and less sensitivity to promoter strength or the cell cycle time.

Our analytical solution to the bistable model is not exact and contains two types of approximations. The first is the interpretation of transcription and translation as bursts of random but instantaneous protein production, as described explicitly in Eq. (2). The second is the use of Eq. (1) and all of the corresponding mathematical approximations used in its derivation. To test the validity of these analytical approximations, we performed Gillespie simulations on the same bistable model (see Methods for more details). For the simulations, we do not interpret protein production as a bursting mechanism but explicitly encode the birth and death reactions of protein and mRNA using p_{RNA}, q_{RNA}, p_{prot}, and q_{prot} to drive protein number fluctuations over time. Cells are initialized in the low state, and we record the mean first passage time to reach the threshold, c. The mathematical approximations used to derive Eq. (1) are also absent in the simulations. However, Fig. 2b shows that the corresponding rates determined from simulations have good agreement with the analytical calculations, providing support

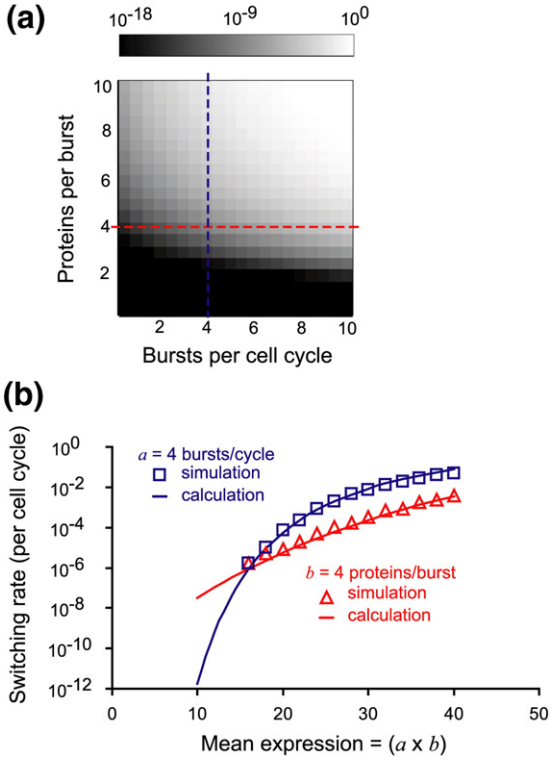

Fig. 2. Burst size and frequency determine the stochastic switching rate. (a) Switching rates determined from Eq. (1) using bursts characterized by the indicated values of the burst frequency and size. (b) Cross section of (a) along the red line, corresponding to a constant translation efficiency, and the blue lines, corresponding to a constant transcription rate. Overlaid in red triangles and blue squares are values determined from stochastic simulations. For the simulations, $q_{RNA}=60$ and $q_{prot}=1$, while p_{RNA} and p_{prot} were tuned to give the indicated values of a and b.

for the validity of the analytical approximations across a range of parameters.

Protein bursts from transcription factor binding fluctuations

The preceding example considered switching by bursts of protein production from unregulated transcription and translation. In many cases, though, the transcription rate is instead regulated by the binding of transcription factors that can serve as repressors or activators. However, even in the presence of such regulation, protein production can still be approximated by our reduced two-parameter description of bursts for certain cases. We will explore the relevant regimes where such an approximation may be valid.

Consider the case of a single copy of a transcription factor, Y, which functions as a repressor for a promoter of interest. Let Y bind to the promoter with a rate k_a and unbind with a rate k_d, to give a binding constant of $K_{eq}=k_a/k_d$. If there are multiple, but constant, molecules of Y, assuming unimolecular dissociation and bimolecular association kinetics, the concentration of Y can be included in the effective association rate, k_a. When the promoter is bound by Y, there is no transcription. When the promoter is clear of Y, there is transcription and translation followed by mRNA and protein degradation with rates of p_{RNA}, p_{prot}, q_{RNA}, and q_{prot} (=1), respectively.

For this system, the average rate of transcription is $p_{RNA}/(1+K_{eq})$, which is the original transcription rate multiplied by the fraction of time that the promoter is not bound by repressor. Then, in a manner similar to before, we can define the burst frequency and burst size as

$$a = p_{RNA} / \left(1 + K_{eq}\right) , \quad b = p_{prot} / q_{RNA} \qquad (3)$$

The interpretation in Eq. (3), however, only holds if we make the equilibrium assumption that binding and unbinding of Y to the promoter occur faster than the rates of transcription and translation. This assumption allows the effect of Y's binding fluctuations to be described by the time-averaged equilibrium constant K_{eq}, instead of the actual kinetic rates. However, it has been suggested that the assumption of fast protein–DNA binding kinetics may not be valid for some systems.[21,22] Thus, we reinterpret Eq. (1) in the context of explicit transcription factor binding fluctuations.

Once the repressor Y dissociates from the promoter in our system, we will assume that there will be an exponentially distributed waiting time with an average of $(k_a)^{-1}$ before Y rebinds. Such exponential kinetics of transcription factor rebinding to an operator have been observed *in vivo* for single cells.[28] However, the rate constant k_a corresponds to the macroscopic observed association rate, and in general, the rebinding time on a microscopic basis will not follow simple exponential kinetics.[29] If the association rate k_a is not diffusion limited, which could occur in the case of competitive binding, allostery, or other modes of binding regulation, the assumption of an exponentially distributed rebinding time $(k_a)^{-1}$ is likely still valid (see Methods). Even in the case of diffusion-limited association, if we only consider microscopic binding fluctuations that lead to productive binding by RNA polymerase, we can interpret an effective, macroscopic binding and dissociation rate of the repressor determined by the frequency of RNA polymerase binding events. In fact, in the case of competitive binding by RNA polymerase and the repressor, rebinding will no longer be diffusion limited, supporting the use of an exponential $(k_a)^{-1}$ time.

Given the assumption above, the number of proteins produced after Y dissociates from the promoter will be proportional to the time that the promoter remains clear, which should also be well described by an exponential distribution. We approximate the resulting burst frequency and burst size based on the following interpretation. The burst frequency is set by the frequency of dissociation, k_d. However, only the fraction of dissociation events that produce at least one transcription event are

observed as bursts, which is given by a comparison of k_a and p_{RNA}. In addition, the rate of dissociation is affected by the probability of the repressor being bound, governed by K_{eq}, since if the repressor is never bound, there will be no new bursts, regardless of k_d. The burst size is determined by the translation rate and mRNA lifetime as before but multiplied by the average number of transcripts produced before the repressor rebinds, given that there must be at least one transcript in order to define a burst. With this approach, the transcription factor binding fluctuations produce proteins in bursts that can be approximated by a burst frequency and burst size of

$$a = \frac{p_{RNA}}{p_{RNA} + k_a} \frac{K_{eq}}{1 + K_{eq}} k_d$$
$$b = (p_{RNA} / k_a + 1)(p_{prot} / q_{RNA}) \quad (4)$$

As should be expected, the mean value of expression, determined by ab, is exactly the same for the out-of-equilibrium case in Eq. (4) and the equilibrium case in Eq. (3). However, the bursting behavior is clearly different. We note that if binding fluctuations are much faster than the transcription rate so that $k_a, k_d \gg p_{RNA}$, the expressions in Eq. (4) reduce to the equilibrium limit in Eq. (3). On the other hand, if binding fluctuations are slow so that $k_a, k_d \ll p_{RNA}$, and $K_{eq} \gg 1$, the expressions reduce to

$$a = k_d, \quad b = (p_{RNA} p_{prot}) / (q_{RNA} k_a) \quad (5)$$

In this last scenario, there is high repression of the promoter by tight binding from Y. The rate-limiting step for expression is the occasional dissociation of Y, with rate k_d, so that the burst frequency is simply k_d.

Role of transcription factor kinetics on switching

The expressions in Eqs. (4) and (5) allow us to interpret protein production within the framework of bursts even in the presence of the kind of transcription factor regulation described above. Thus, we can again apply our previous solution to threshold crossing by bursting protein production in Eq. (1), but with a different interpretation of the bursting parameters a and b. Previously, in Fig. 2, we explored the dependence of switching rate on protein number fluctuations determined by the promoter strength and translational efficiency. Our example now explicitly includes transcription factor kinetics as an additional mechanistic source of protein number fluctuations; hence, we examine how changing the transcription factor kinetics alters the effective burst size and frequency, which, in turn, alters the switching rate.

To explore how the absolute transcription factor kinetics, rather than the absolute promoter strength, can affect the switching rate, we fix K_{eq} to be constant and change k_d and $k_a = k_d K_{eq}$, accordingly. All other parameters are fixed. This changes the absolute rate of binding fluctuations, without changing their equilibrium properties. In the absence of positive feedback, the mean value of protein expression would remain the same, but the underlying mechanism of protein production and, hence, fluctuations would be altered. We examine how such promoter changes affect the rate of crossing a threshold.

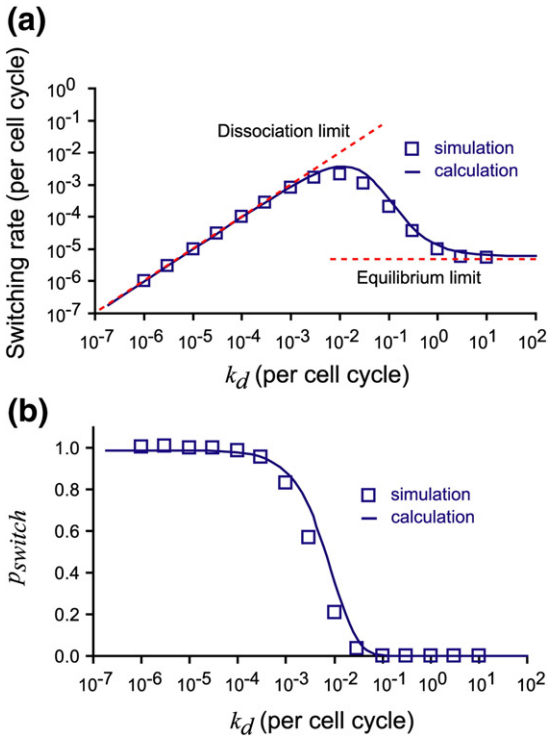

Fig. 3. Transcription factor kinetics determine switching rates in a non-monotonic manner. (a) Switching rate as a function of the transcription factor dissociation rate using Eqs. (1) and (4). Transcription factor kinetics are changed while keeping the equilibrium binding constant fixed. The non-monotonic dependence of the switching rate on k_d is a result of the limiting behaviors plotted as red broken lines. Overlaid in blue squares are values determined from stochastic simulations. For the simulations, $c = 100$, $p_{RNA} = 10$, $p_{prot} = 600$, $q_{RNA} = 60$, and $q_{prot} = 1$, while k_d and k_a were changed, keeping $K_{eq} = k_a/k_d$ constant. (b) The probability of a single dissociation event leading to a switching event is plotted as a function of the transcription factor dissociation rate for the parameters from (a). Rare dissociation events likely to cause switching correspond to $p_{switch} \sim 1$ and a single-step switching mechanism. Frequent dissociation events unlikely to cause switching correspond to $p_{switch} \sim 0$ and a diffusive switching mechanism. Overlaid in blue squares are values determined from stochastic simulations (see Methods). Standard errors for the simulation results are in the range of 1–3%, which are much smaller than the size of the data point on the logarithmic scale.

The calculated and simulated results of Fig. 3a show that the switching rate has a non-monotonic dependence on the transcription factor kinetics. We first consider the limits of slow and fast transcription factor kinetics. Clearly, the switching rate cannot be faster than the repressor dissociation rate, k_d, since no protein expression can occur until a dissociation event occurs. Thus, as k_d approaches zero, we also expect the switching rate to approach zero. On the

other hand, as the dissociation rate, k_d, increases, while keeping K_{eq} constant, the switching rate should converge to the equilibrium solution obtained by using the parameters in Eq. (3).

The maximum switching rate occurs for intermediate transcription factor kinetics, so that the equilibrium solution always underestimates the possible switching rate. This can be understood from the large protein number fluctuations occurring in the presence of slow transcription factor kinetics, as has been discussed previously in the context of slow toggling of a promoter.[30,31] Holding K_{eq} constant keeps the mean protein numbers across the population fixed for all k_d, but as k_d decreases, the protein distribution becomes more disperse, which increases the rate of reaching a high threshold. Interpretations of experiments in the context of the non-monotonic relation are discussed in later sections.

Switching probability of a single transcription factor dissociation event

The non-monotonic plot in Fig. 3a implies that two systems with the same average promoter properties and switching rates can have very different underlying mechanisms. In the last section, we explained the non-monotonic behavior by considering the mathematical limits of the underlying equations. However, we also seek to formulate a concise, conceptual description of the different parameter regimes. Furthermore, such a description should provide a way to distinguish between two different regulatory systems with the same switching rate.

We start with a qualitative description of how fast or slow transcription factor kinetics can provide the same switching rate. In the case of faster transcription factor kinetics, the bursts are small but frequent, and the accumulation of many small bursts can occasionally reach the threshold necessary for switching. This is similar to a diffusive random walk in a potential that approaches a critical threshold through many small steps. In the case of slow transcription factor kinetics, bursts are very large but also very rare. However, when a burst occurs, it is very likely that there will be enough proteins to reach the threshold. This corresponds to a one-step switching process.

One way to quantify the conceptual difference between these two molecular scenarios is to consider the number of molecular steps that occur for each case. In particular, we examine the switching probability of a single transcription factor dissociation event. Assume that the repressor Y binds the promoter and then dissociates. The probability that this particular dissociation event leads to switching can be calculated by considering the simple relation for $K_{eq} \gg 1$:

$$k = k_d p_{switch} \tag{6}$$

Here, p_{switch} is the probability that any single dissociation event will lead to switching and provides a basis for describing the two molecular scenarios. We calculate p_{switch} directly from the original parameters used in Fig. 3a by substituting into Eq. (6) [see Eq. (21) in Methods for an explicit form], and we plot the results in Fig. 3b along with the corresponding simulation results that determine p_{switch} through direct counting. Figure 3b shows that as k_d increases, $p_{switch} \sim 0$, while as k_d decreases, $p_{switch} \sim 1$. These two p_{switch} regimes correspond to switching by diffusive random walks or a single rate-limiting step, as illustrated in Fig. 4. In Eq. (6),

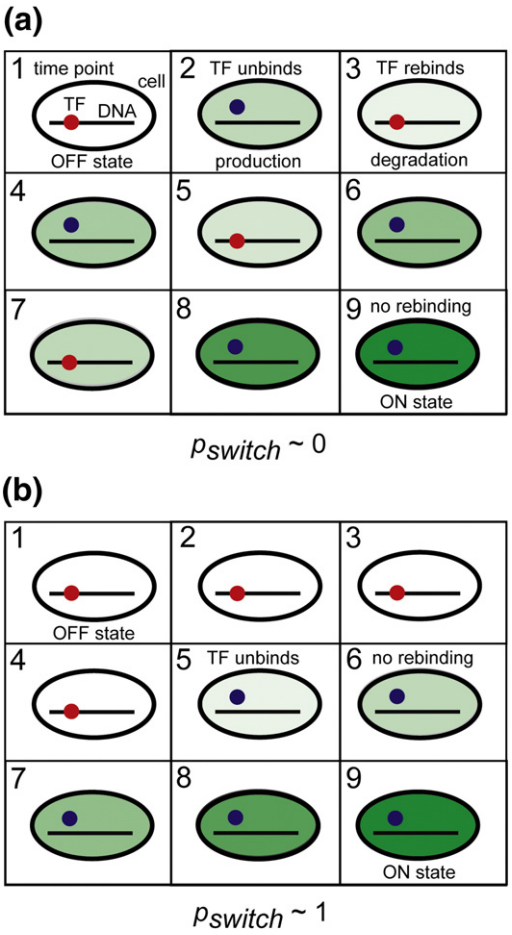

Fig. 4. Two contrasting mechanisms can generate stochastic switching. (a) One mechanism of stochastic switching requires many molecular fluctuations. For example, a transcription factor regulating gene expression may bind and unbind many times. The probability of any individual unbinding event leading to a change in phenotype is very small, which we classify as $p_{switch} \sim 0$. In this case, the protein numbers may fluctuate over time, similar to a diffusive, random walk. When the protein number is large enough to activate a feedback mechanism, the cell remains in a state of high expression. In this figure, the promoter state is indicated by the color and position of the TF, and the color of the cell body indicates the level of protein expression. Each panel indicates a subsequent point in a time series. (b) An alternative mechanism of stochastic switching relies on the rare occurrence of specific molecular fluctuations. A transcription factor or complex may only dissociate rarely. However, each dissociation event has a high probability of leading to a change in phenotype. We describe this scenario as $p_{switch} \sim 1$. In this case, a single molecular step is responsible for stochastic switching.

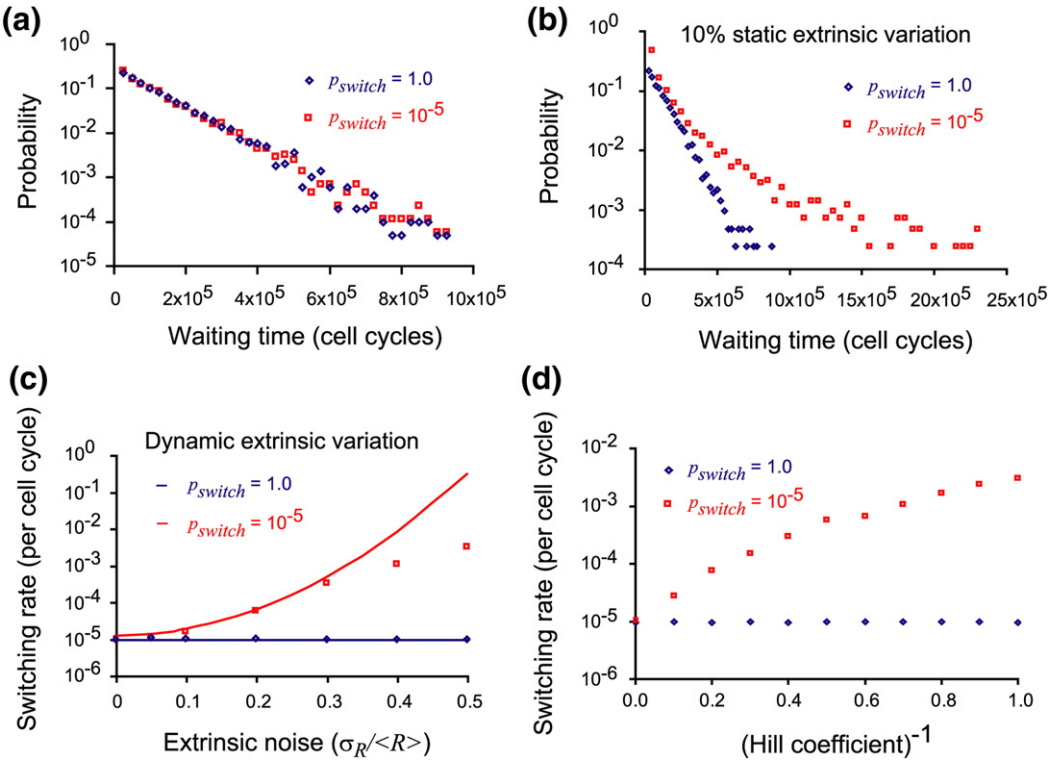

Fig. 5. Robustness of switching rates to variation in parameters. (a) The waiting time distribution for two sets of parameters with identical switching rates and identical average expression levels is determined from stochastic simulations for the same set of parameters as in Fig. 3a. The blue diamonds correspond to the case with $k_d = 10^{-5}$ and $p_{switch} \sim 1$, while the red squares correspond to the case of $k_d = 1$ and $p_{switch} = 10^{-5}$. These very different mechanisms have identical kinetic properties. (b) In the presence of static extrinsic variation in cellular parameters, the waiting time distribution for switching by the $p_{switch} = 10^{-5}$ case no longer follows simple exponential kinetics. The same average parameter values are used as in (a). (c) The response of the switching rate to extrinsic noise, or variation in its parameters, is shown using the same conditions as in (a), as calculated by Eq. (8) (lines). The switching rate for $p_{switch} \sim 1$ is robust to parameter fluctuation, while the $p_{switch} \sim 0$ is not. Simulation values are overlaid (squares and diamonds). (d) The switching rate is determined for the same conditions as in Fig. 4a, but with a positive feedback function showing different degrees of cooperativity. The switching rate for $p_{switch} \sim 1$ is insensitive to the precise form of the positive feedback, but the $p_{switch} \sim 0$ shows a strong dependence.

p_{switch} is not independent of k_d; thus, there are at most two pairs of k_d and p_{switch} possible for a specific k with other parameters fixed.

We will use $p_{switch} \sim 0$ and $p_{switch} \sim 1$ to label the two limiting mechanistic scenarios for stochastic switching. Thus, the initial observation is that it is not possible to determine whether $p_{switch} \sim 0$ or $p_{switch} \sim 1$ for a system from only observing the mean switching rate, k. We further show that there is no difference in the kinetic properties of the two cases by characterizing the waiting time distributions. We select two sets of parameters with identical K_{eq} and k but different k_d values corresponding to $p_{switch} \sim 0$ and $p_{switch} \sim 1$. Simulations were performed in the same manner as for Fig. 3 by explicitly simulating every microscopic reaction, with no interpretative simplification of the protein production process. However, instead of reporting only the average time to reach the critical threshold, we recorded the entire distribution of times for many iterations of the simulation, which would correspond to measurements on different cells from a population. Figure 5a shows simulation results for both the $p_{switch} \sim 0$ and $p_{switch} \sim 1$ processes, which share identical, exponentially distributed waiting times. We conclude that both scenarios exhibit, to a good approximation, identical switching rates and even identical kinetic distributions.

Robustness of switching rates to parameter fluctuations

We further explored the macroscopic differences between the $p_{switch} \sim 0$ or $p_{switch} \sim 1$ mechanisms, since the basic kinetic properties appear identical. Although the molecular mechanism of $p_{switch} \sim 0$ or $p_{switch} \sim 1$ cannot be determined from the switching rate alone, we find that the two cases exhibit very different sensitivities to perturbations on system parameters. We will use the simplified expressions in Eq. (5) to illustrate this. Upon substituting Eq. (5) into Eq. (1), we obtain

$$k = (aR)^a \frac{e^{-aR}}{\Gamma(a)} = (k_d R)^{k_d} \frac{e^{-k_d R}}{\Gamma(k_d)} \qquad (7)$$

where $R = c/ab = (cq_{RNA}K_{eq})/(p_{RNA}p_{prot})$. The variable R contains all of the parameters that define the

system other than the transcription factor fluctuation rate, k_d. Assume that R is not fixed but follows a normal distribution with mean $\langle R \rangle$ and variance σ_R^2. For example, heterogeneity between cells could arise from different RNA polymerase or ribosome numbers, which would affect p_{RNA} and p_{prot}, or other global factors such as cell size and ATP concentration that similarly influence R. The transcription factor concentration, which alters the effective k_a and K_{eq}, is also included in R. Such variation in the parameters of a gene network, termed *extrinsic noise*, has been observed to be significant in previous studies.[17,18,32,33] We consider how switching properties change as a function of the magnitude and rate of variation in R.

If the fluctuation rate of R is slower than the rate of switching, the distribution of R will change during the switching process and not remain normal. For example, we can consider a case where extrinsic noise arises from purely static heterogeneity between cells. Then, the switching process no longer needs to follow exponential kinetics. A rapidly switching subpopulation with small R dominates the switching kinetics on short time scales, but after long times, a subpopulation with large R determines the switching rate. In Fig. 5b, upon adding 10% static variation to R for the same simulation parameters as in Fig. 5a, we observe that the $p_{switch} \sim 1$ waiting time distribution appears to remain exponential. However, the presence of static extrinsic noise results in a nonexponential waiting time distribution for $p_{switch} \sim 0$; hence, in this case, we do not define a single rate constant.

If the fluctuation rate of R is faster than the rate of switching, during the process of switching, we can assume that the distribution of R remains constant. In this case, we can define an average rate constant to describe the switching process because the switching kinetics are the same for both short and long time scales. We average Eq. (7) over a normal distribution of R and approximate the switching rate in the presence of rapidly changing cell-to-cell variation (see Methods) as

$$\langle k(R) \rangle \approx k(\langle R \rangle) e^{(a\sigma_R)^2/2} \tag{8}$$

From Eq. (8), we find that the switching rate is more than exponentially sensitive to the magnitude of variation in the system parameters summarized by R. However, the sensitivity to these parameter fluctuations is controlled by $a \approx k_d$ for k_a, $k_d \ll p_{RNA}$, and $K_{eq} \gg 1$. For a given switching rate k, if there are two possible values of k_d and p_{switch}, the larger value of p_{switch} is always associated with the smaller value of k_d. Thus, the mechanistic scenario of $p_{switch} \sim 1$, corresponding to switching by a single molecular event, results in a more robust switching rate. In the limit of very small k_d,

$$\langle k(R) \rangle \approx k_d(1 - k_d \langle R \rangle) \tag{9}$$

The linear dependence of k on R for small k_d means that positive and negative fluctuations of R will cancel, and the average switching rate is almost independent of variation in parameters. In contrast, a mechanism with $p_{switch} \sim 0$, corresponding to switching by many molecular steps, does not give a robust switching rate. In the limit of very large k_d, the behavior of the exponential term in Eq. (8) dominates, and the switching is determined by the size of the variation in R. Cells with small R would switch rapidly. If the rate of R fluctuations is faster than the fastest switching rate, as we have assumed, then the fluctuations in R would ensure that some fraction of unswitched cells will always have small R. Then, fast switching for cells with small R would dominate the entire switching process. Figure 5c shows the contrasting sensitivities of the $p_{switch} \sim 0$ and $p_{switch} \sim 1$ cases to such variation in R, using parameters yielding identical switching rates when $\sigma_R = 0$.

Dependence of switching rates on positive feedback function

The robustness of the $p_{switch} \sim 1$ case implies that the simplicity of the model used to derive Eq. (1) may not limit its applicability to some real biological networks. For example, the form of the positive feedback we have been using in our mode, as stated previously, is considered to be a step function, with burst frequency a for $X < c$ and burst frequency $a*$ for $X > c$. We can use a more general form of positive feedback, such as a Hill equation of the form frequency $= a + (a^* - a)X^n/(X^n + c^n)$. Our step function corresponds to the Hill coefficient $n \to \infty$. We can consider more realistic cases such as $n = 2$ or $n = 3$, which are plotted in Fig. 5d. We performed simulations using this positive feedback function, initializing cells at the mean expression in the low state and calculating the average time to reach the threshold. For these simulations, the transcription rate is recalculated after every change in protein number using the Hill equation with a constant threshold value.

The $p_{switch} \sim 1$ mechanism is relatively insensitive to the Hill coefficient or the particular form of the positive feedback. Because $p_{switch} \sim 1$ corresponds to rare, large bursts of protein expression, just before a cell switches, it will still have very low protein numbers and have a burst frequency close to a. Once a single burst occurs, the protein number will reach the threshold c with high probability because the burst is very large. The precise form of the positive feedback is irrelevant to the switching event because only the first large burst is necessary for switching.

On the other hand, for the $p_{switch} \sim 0$ case, the form of the positive feedback is very relevant, since reaching the threshold occurs through many burst events. The timing of each burst depends on the form of the positive feedback, and the exact sum of these many bursts is important for determining the switching rate. Thus, as the Hill coefficient decreases, the switching rate increases, since the feedback becomes more important even at low expression. The simplified model using a stepwise threshold is inaccurate for this case, and at extreme

limits, such as $n=1$, the interpretation of bistable switching may also fail, since nonlinear positive feedback is no longer present.

Analysis of bursting and switching in the *lac* operon

We now analyze switching for a model experimental system, lactose metabolism in *Escherichia coli*, as threshold crossing by bursting protein production. In this system, active transport of a small-molecule inducer by the lactose permease provides a positive feedback mechanism; thus, under certain growth conditions, cells will express either very low or very large amounts of permease.[4] A recent experimental work characterized the permease production and threshold for switching from the uninduced to induced state of the *lac* operon.[13] This work proposed that two distinct physical processes, partial and complete dissociation of the tetrameric repressor from the promoter, generated "small" bursts and "large" bursts of permease expression, respectively. Operator site mutations showed that the specific molecular event of a complete dissociation was likely to be the rate-limiting step for reaching the switching threshold. We examine these claims by estimating switching rates from reported burst and threshold data, treating small and large bursts as distinct classes of events with unique burst properties.

We note that evaluating experimental data, even for a model system such as the *lac* operon, is difficult for several reasons. The main problem is that many of the required parameters are poorly characterized. Direct measurements of real-time protein production dynamics, rather than model-dependent inference from static population measurements, are rare. In addition, single-cell protein expression levels are often reported in arbitrary fluorescence units, rather than absolute molecule numbers, making it difficult to compare different data sets. Furthermore, bistability of the *lac* operon is highly dependent on growth conditions, including glucose and inducer concentration.[34] Thus, the proper experimental data must be assembled for any calculation. For these reasons, we rely on a primary reference,[13] which reports a set of absolute, real-time data, and make comparisons with other references as appropriate (see Methods for further details on how experimental data were applied).

Fluctuations of permease numbers in uninduced cells have been reported to be the result of small bursts, which have $a_{\mathrm{small}}\sim 0.5$ and $b_{\mathrm{small}}\sim 2.5$.[13] Similar values of the burst frequency and burst size in the absence of inducer have been previously reported for the repressed *lac* promoter, and small bursts are interpreted as the number of proteins translated per transcript, with only one transcript being produced at a time.[35,36] Only one study has made a direct measurement of the functional threshold for induction, which is reported to be 375 permease molecules per cell for a specific growth condition.[13] However, it is possible to set bounds on the threshold from other experimental or simulation data, by considering the range of expression for uninduced and induced cells. Clearly, the threshold should lie in between these two populations' expression levels. This range of excluded expression levels in bimodal populations, where the probability of occurrence is small, is estimated to lie in the range of approximately 10–30%[34] of the maximal expression level in other literature. If fully induced cells can have up to ~1200 permease molecules per cell,[13] $c=375$ would be consistent with these expectations.

If we take $c=375$, we can test whether the frequent small bursts of expression in uninduced cell are responsible for switching. The predicted switching rate, if the small bursts are the mechanistically relevant fluctuation, would be $k_{\mathrm{small}}\sim 10^{-63}$. The calculated rate is unrealistically low, even if the values of a_{small}, b_{small}, or c are changed by an order of magnitude, given that the observed switching rate is 0.004 for 40 μM of the inducer TMG.[13] Thus, the frequent and small fluctuations observed in uninduced cells are not likely to drive switching.

Large bursts are suggested to arise from multiple rounds of transcription following complete dissociation of the repressor and the subsequent slow rebinding in the presence of inducer. Experimentally, large bursts are difficult to characterize because of their rarity. However, in the presence of 200 μM TMG, with the permease feedback removed, the frequency of large bursts was reported to be more than 10 times smaller than the frequency of small bursts.[13] At lower concentrations, large bursts were too rare to observe directly, while the frequency of small bursts remained similar.[13] Thus, we set a bound of $a_{\mathrm{large}}<0.05$ for the wild-type system at 40 μM TMG. Given that the switching rate is 0.004, we also know that $a_{\mathrm{large}}>0.004$. We calculated the required b_{large} to give the correct switching rate of 0.004 for this a_{large} range and show the results in Fig. 6. Predicted burst sizes are all possible, since ~1000 molecules per cell cycle is a biologically feasible permease expression rate. The necessary burst sizes have further implications on the molecular mechanism of large bursts, which we discuss in the Appendix. The corresponding p_{switch} values also suggest that the complete transcription factor dissociation occurs less than ~10 times before a switching event occurs. As a comparison, obtaining a switching rate of 0.004 using any biologically reasonable parameters ($b\geq 1$ molecule, $a<1000$ per cycle) can yield up to 10^5 binding fluctuations per switching event.

Based on these results, we conclude that estimation of switching rates from small and large burst properties is consistent with the previous qualitative interpretation that rare large bursts, rather than frequent small bursts, drive switching of the *lac* operon with high probability. Importantly, we find that applying our model to experimental data generates biologically feasible results, providing a test of its accuracy and applicability. Experimental values necessary for these kinds of calculations are

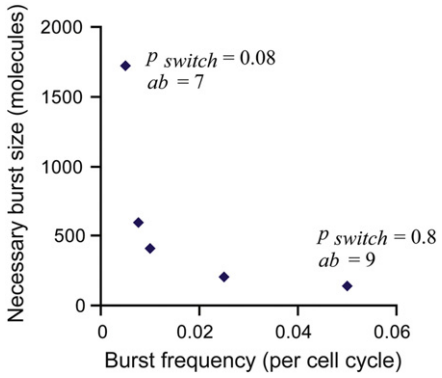

Fig. 6. Range of parameters with correct switching rate for *lac* operon. The switching rate of the *lac* operon in the presence of 40 μM TMG is 0.004 per cell cycle. Experimental measurements set limits of $0.004 < a_{large} < 0.05$. The corresponding value of b_{large} necessary to generate a switching rate of 0.004, given an experimentally measured threshold of 375 molecules, is shown in the plot. The values of p_{switch} and $a_{large}b_{large}$ are shown for the end points.

sparse, but as real-time single-cell imaging techniques improve, it may be possible to test mechanisms of other genetic switches by direct, simplified calculations.

Discussion

Intrinsic *versus* extrinsic control of switching

The proposal that protein noise, coupled with positive feedback, can result in stochastic switching between bistable states has been discussed previously in the literature.[16] However, the precise nature of the noise is often overlooked or simplified. We show that properties of protein noise, resulting from specific molecular mechanisms such as transcription, translation, or transcription factor regulation, have significant effects on both the rate and the robustness of stochastic switching. In particular, the characteristic frequency and size of protein number fluctuations alter switching properties.

For the case of regulation by a repressor, rapid transcription factor dissociation and rebinding will result in small, frequent bursts of protein expression and many unbinding events before a switching event occurs. A low probability of switching for each molecular event, which we describe by $p_{switch} \sim 0$, however, results in a mechanism of switching that is very sensitive to parameter fluctuations. The effect of extrinsic noise, or changes in these parameters, can dominate the rate of switching, as shown by Eq. (8). In this case, the environment or global fluctuations of a cell control the rate of switching, which we term *extrinsic control*.

When protein bursts are large but infrequent, any dissociation event by a repressor has a high probability of resulting in switching. Thus, each molecular event is important, which we describe by $p_{switch} \sim 1$. This molecular mechanism is robust to extrinsic noise, and the switching rate is controlled by the average parameters of a cell. In such a case, the switching rate is determined only by the intrinsic biological parameters encoded in the genome, which we term *intrinsic control*.

Recently, there have been a number of studies exploring the optimal rate of stochastic switching between phenotypes in order to maximize fitness.[37,38] One of their conclusions is that a species may evolve an optimal switching rate if there are environmental fluctuations with a predictable time scale. The case of $p_{switch} \sim 1$ with intrinsic control could, in principle, allow robust, genetic encoding of a switching rate, but the case of $p_{switch} \sim 0$ with extrinsic control would make it difficult to evolve a specific switching rate. Most mechanistic descriptions of stochastic switching, however, have used generic protein noise implicitly associated with the $p_{switch} \sim 0$ case, which is unlikely to result in the evolution of robust switching time scales. We show, though, by considering differing noise mechanisms that the $p_{switch} \sim 1$ case may be more applicable to the scenario of specifically evolved rates.

The instability of the $p_{switch} \sim 0$ case suggests that this class of biological switches should also be very difficult to model quantitatively. Small perturbations in the inputs cause large variations in the switching rate; hence, an accurate model would require exact values for all of the biological parameters. However, many of these parameters are difficult to measure *in vivo* and can only be estimated. Even if the parameters are measured, most assays determine the average value, such as the average number of transcription factors or average cell size. However, Eq. (8) shows that the average switching rate is not simply the switching rate using the average value of the parameters. In the $p_{switch} \sim 0$ case, the variation in the parameters, which is rarely measured, can be even more important than the actual average values.

Efficiency of $p_{switch} \sim 1$ mechanism

We discussed the functional difference between intrinsic and extrinsic control mainly with respect to robustness, but other functional properties are possible as well. In particular, the two classes of switching use different amounts of total protein to incur the same switching rate, which may correlate with different metabolic costs. If switching occurs through the action of bursts with frequency a and size b, we can calculate the expected mean expression level of uninduced cells to be $(1-p_{switch})ab$. In the case of $p_{switch} \sim 0$, many "failed" bursts would accumulate in the population of uninduced cells, affecting the expression level of uninduced cells. However, for $p_{switch} \sim 1$, cells that experience a large burst always leave the population of uninduced cells; hence, uninduced cells have almost no protein expression.

Thus, another functional consequence of the $p_{switch} \sim 1$ mechanism is a minimal protein level for

uninduced cells regardless of switching rate. This feature could be advantageous for avoiding the metabolic cost of synthesizing proteins that provide no phenotypic benefit. For the example of the *lac* operon, in cells that switch to the induced state, there is an increased metabolic cost because of high protein expression, but there is also a compensating benefit of lactose metabolism, a topic analyzed previously.[39] Uninduced cells, by definition, lack enough proteins to provide the benefit of lactose metabolism. In the hypothetical case that $p_{\text{switch}} \ll 1$, bursts that fail to drive switching would have a metabolic cost for uninduced cells, but these uninduced cells would still be unable to survive on lactose as a carbon source. A $p_{\text{switch}} \sim 1$ mechanism, though, guarantees that every lactose metabolism protein produced is found in a cell that engages in active lactose metabolism. In this way, $p_{\text{switch}} \sim 1$ would be efficient.

Therefore, the expression level in uninduced cells may relate to the value of p_{switch}. Experimentally, it was observed that the permease numbers in the uninduced population are extremely small and relatively insensitive to inducer concentration. Single-molecule measurements indicate mean expression levels of 0.9 ± 0.1 and 1.2 ± 0.1 molecules per cell, respectively, for 0 and 40 μM TMG.[13] However, the switching rate changes dramatically as a function of inducer concentration, implying that the uninduced population expression is insensitive to any possible changes in a_{large} and b_{large}. Quantitatively, even though $a_{\text{large}}b_{\text{large}} \sim 5$–$10$ molecules per cell for the range of possible values in Fig. 6, the mean expression of uninduced cells can remain unchanged at one molecule per cell, as observed, if $p_{\text{switch}} \sim 1$. Thus, the extremely low level of expression in the uninduced cells of the *lac* operon, regardless of switching rate, suggests efficient conservation of protein expression. This second approach to analyzing experimental data supports the idea that $p_{\text{switch}} \sim 1$ and also provides further insight into a potential feature of the $p_{\text{switch}} \sim 1$ mechanism.

Classifying p_{switch} of biological systems

We provided estimates of p_{switch} for the *lac* system in preceding sections using experimental data. Experimental measurements necessary to calculate p_{switch} directly for other systems have not been reported. However, even if precise values of burst and threshold values are not available, we can make qualitative predictions about whether the system is closer to $p_{\text{switch}} \sim 0$ or $p_{\text{switch}} \sim 1$ mechanisms based on properties of the switching behavior.

The non-monotonic behavior of the switching with respect to burst size and frequency in Fig. 3a provides a key qualitative predictor of whether $p_{\text{switch}} \sim 0$ or $p_{\text{switch}} \sim 1$. For the case of $p_{\text{switch}} \sim 1$, a specific molecular reaction is rate-limiting, and increasing the rate of that reaction increases the switching rate. This was demonstrated for switching of the *lac* operon by changing the rate of complete repressor dissociations, which lead to an increase in the rate of switching.[13] In lactose metabolism, as well as other stochastic bacterial systems such as lambda phage lysogeny, transcription factors are involved in multimeric binding to DNA, resulting in DNA loops. DNA looping can create a very small effective rate of complete dissociation, allowing slow transcription factor kinetic to arise from this effective dissociation rate of the entire promoter complex. Such all-or-none dissociation has been previously proposed to explain robust switching rates.[24]

While robust switching rates with $p_{\text{switch}} \sim 1$ may be hypothesized to have a fitness advantage for certain systems, others may benefit from extrinsic control. For example, some responses, such as potentially mutagenic DNA uptake or competence may be more beneficial in the presence of environments characterized by uncertainty or large global fluctuations. Stochastic switching of *B. subtilis* competency is a well-studied model system. A recent work studied the effect of noise on switching by increasing the rate of transcription while reducing the rate of translation for the regulatory protein, ComK.[12] They found that the rate of competency was decreased by a factor of 15. If the frequency of bursts is increased while keeping the mean expression constant and the resulting switching rate decreases, the situation would correspond to the right side of the non-monotonic plot in Fig. 3a with $p_{\text{switch}} \sim 0$.

It appears that systems close to both the $p_{\text{switch}} \sim 0$ and $p_{\text{switch}} \sim 1$ cases are biologically relevant. The two cases can be distinguished by the relationship between burst frequency and size with the switching rate. Another test would be to measure the response of the switching rate to small perturbations in the system parameters. An exponentially sensitive response indicates that $p_{\text{switch}} \sim 0$, while a more robust response indicates that $p_{\text{switch}} \sim 1$. Finally, an even more direct measurement of p_{switch} would involve determination of the switching rate and the dissociation rates, in conjunction with Eq. (6). Measuring of dissociation rates *in vivo*, though, is not straightforward. Furthermore, the effective dissociation rate of promoter complexes and looped structures is more difficult to determine, even when extrapolating from *in vitro* measurements.

What we have shown, though, is that the understanding of stochastic biological switches requires a mechanistic model of molecular fluctuations and outlined an approach for understanding the relationship between fluctuation mechanisms and switching rates for several examples. While the methods in this work are probably most relevant to single-cell prokaryotes, extensions to other systems may be possible if the complexity of underlying protein dynamics and promoter states can be approximated in a succinct manner, particularly with analytical expression.[40,41] Although we deal primarily with repressor unbinding, other examples of a promoter slowly switching between two states for other reasons, such as activator binding,

methylation, or chromatin remodeling, may be applicable using similar approaches. The resulting mechanistic understanding of such stochastic switching may not only help explore natural phenotype switching but also provide design principles for the construction of robust synthetic switches.

Methods

Simulations

Switching rates were determined using Gillespie simulations[42] with reactions as described in Fig. 1. Parameters for simulations have been described in the main text. Simulations were initialized at the mean expression level ab. For Fig. 2, equilibrium distributions were obtained by 100,000 simulation steps before the initial zero time point. For Fig. 3, the system was initialized with the transcription factor bound to the DNA. Simulations were repeated for a minimum of 1000 times in order to determine the average switching rate, with most repeated over 4000 times. As many as 10^8 steps were necessary to evaluate the switching rate directly. For Fig. 5a and b, simulations were repeated a minimum of 4000 times to determine waiting time distributions. A threshold parameter was drawn from a random normal distribution for each run of the simulation used to construct Fig. 5b. For Fig. 5c, simulations were repeated a minimum of 300 times for fixed values of the threshold parameter to determine the switching rate across a range of thresholds. Then, a normally distributed weighting was applied to the switching rates to calculate the average switching rate resulting from a normal distribution of thresholds with the indicated variance.

Calculation of switching rates

We calculate the rate of threshold crossing by the production of proteins in bursts. Assume that the production of protein X occurs in uncorrelated, instantaneous bursts, as described in the main text. The burst frequency is a if $X<c$, and the burst frequency is a^* if $X\geq c$, with $a<a^*$. Assume that the states are well separated so that $ab\ll c\ll a^*b$. Otherwise, bistability would not be observed. In the absence of any positive feedback, the steady-state distribution of proteins per cell, $P(X)$, is

$$P(X) = \frac{X^{a-1}e^{-X/b}}{\Gamma(a)b^a} \tag{10}$$

when solved with continuous variables.[27] We use this analytical framework as a basis because of its simple and extendable burst interpretation, although other models of bursting expression are possible as well.[30,31] Assume that the system starts with all cells near $X=ab$. Bistability requires that the transition rate to $X\geq c$ be slower than the time scales of protein production and degradation. In other words, we can assume that the distribution near $X=ab$ even in the presence of positive feedback is similar to Eq. (10) if the system is in a quasi-steady state. Then, for each burst, we calculate the probability of that burst crossing the threshold c if the initial state is $X=x$. Because of the quasi-steady-state assumption, we only need to consider the probability for single bursts and not the cumulative sum of bursts, since the static distribution of X already captures the dynamics of failed threshold crossing.

The probability that a single burst with an exponentially distributed size is large enough to cross the threshold, starting from an initial condition of $X=x$, is

$$F(x)\equiv\int_{c-x}^{\infty}\frac{e^{-X/b}}{b}\mathrm{d}X = e^{-(c-x)/b} \tag{11}$$

Assuming that the occurrence of bursts follows Poisson statistics, if there is an average of a bursts per cell cycle, the probability in one cell cycle of at least one successful event is:

$$p = 1-e^{-aF(x)} \tag{12}$$

The switching rate can now be calculated by averaging Eq. (12) over the distribution of x, which is performed by approximating a discrete summation with a continuous integral.

$$k\approx\int_0^c\left(1-e^{-aF(x)}\right)\frac{x^{a-1}e^{-x/b}}{\Gamma(a)b^a}\mathrm{d}x \tag{13}$$

By substituting $y=(c-x)/b$ and expanding the integrand, we obtain

$$\begin{aligned}k\approx{}&\frac{1}{\Gamma(a)b^{a-1}}\int_0^{c/b}e^{-c/b}(c-yb)^{a-1}e^y\\&\times\left(ae^{-y}-\frac{(ae^{-y})^2}{2}+O\left((ae^{-y})^3\right)\right)\mathrm{d}y\end{aligned} \tag{14}$$

We now simplify this expression with approximations by considering the magnitude of the integrand terms in different parameter ranges. For $a\leq 1$, it is clear that we only need to keep the first of the three terms on the right side of the integrand. If $a>1$, the first of the three terms is also sufficient to evaluate the expression if $y\geq a$. The integral is carried out, however, for $0<y<c/b$. If we apply the same approximation when $y<a$, we will be overestimating the value of the expression on the right by at most a factor of a. We also note that integral of the weighting function $(c-yb)^{a-1}$ carried out for $0<y<a$ will contribute to about ab/c of the total integral from $0<y<c/b$. Since $ab\ll c$, we conclude that the contribution of the whole integrand for $y<a$ is a small fraction of the total integral. Making the assumption that only the first term is necessary for the entire integral will lead to an error of about $a^2b/c\sim O(1)$ for realistic values of a satisfying $ab\ll c$. Thus, we apply this approximation in order to compute the compact solution in Eq. (1).

Calculation of switching rate in the presence of extrinsic noise

In the main text, we parameterized all of the switch properties, except the repressor unbinding rate, by the variable R. In the presence of extrinsic noise, different cells will have different values for R; hence, we calculate the resulting average switching rate for a large population of cells. We only calculate this result when the extrinsic fluctuations of R are faster than the time scale of switching. Otherwise, the population switching kinetics exhibit a nonexponential waiting time distribution (see the main text); thus, the definition of a switching rate is unclear.

Beginning with the form of Eq. (7), assume that R is subject to a normal distribution with variance σ_R^2. We

assume that the correlation time of R is longer than one cell cycle but shorter than the time scale of the switching rate k. Because the correlation time of R is shorter than the switching rate k, the distribution of R will remain effectively constant over time. Thus, we can approximate the population's switching rate by averaging over a population of cells with different R values as:

$$\langle k(R)\rangle = \int (k_d(\langle R\rangle + \delta))^{k_d} \frac{e^{-k_d(\langle R\rangle + \delta)}}{\Gamma(k_d)} \frac{e^{-\delta^2/2\sigma_R{}^2}}{\sqrt{2\pi}} d\delta \quad (15)$$

To approximate this integral, we consider how the normal distribution minimizes the integrand at large values of δ and instead calculate the integral only near the maximal value of the integrand. We rewrite the integrand in the form $e^{-G(\delta)}$ and find the minimum value of $G(\delta)$ to be at:

$$\delta^* = \frac{-\langle R\rangle - \sigma_R{}^2 k_d + \sqrt{(\langle R\rangle - \sigma_R{}^2 k_d)^2 + 4\sigma_R{}^2 k_d}}{2} \quad (16)$$

This solution satisfies $\delta^* > -\langle R\rangle$, which is required for nonzero values of R. We approximate the decay of $e^{-G(\delta)}$ near δ* by:

$$\delta - \delta^* \approx \sqrt{\frac{-2}{G''(\delta^*)}} = \sqrt{\frac{2}{k_d/(\langle R\rangle + \delta^*)^2 + 1/\sigma_R{}^2}} \quad (17)$$

which leads to the expression

$$\langle k(R)\rangle \approx (k_d(\langle R\rangle + \delta^*))^{k_d} \frac{e^{-k_d(\langle R\rangle + \delta^*)}}{\Gamma(k_d)} \frac{e^{-\delta^{*2}/2\sigma_R{}^2}}{\sqrt{2\pi}} \times \sqrt{\frac{2}{k_d/(\langle R\rangle + \delta^*)^2 + 1/\sigma_R{}^2}} \quad (18)$$

From Eq. (18), we can further simplify in the case that $\sigma_R^2 k_d \ll \langle R\rangle$. We also note the original definition of $R = c/ab \gg 1$, because we required $c \gg ab$ for bistability. Thus, $\langle R\rangle \gg 1$ as well. Then, by Taylor expanding for the value of δ* to first order in $\sigma_R^2 k_d$, we obtain $\delta^* \approx \sigma_R^2 k_d$ and

$$\langle k(R)\rangle \approx k(\langle R\rangle) e^{(k_d\sigma_R)^2/2} \quad (19)$$

We can perform a similar approximation for larger $\sigma_R^2 k_d \sim \langle R\rangle$ to obtain:

$$\langle k(R)\rangle \approx k(\langle R\rangle)(k_d\sigma_R{}^2)^{-k_d/2} e^{(k_d\sigma_R)^2/2} \quad (20)$$

As σ_R approaches $\langle R\rangle$, the rate of increase of $\langle k(R)\rangle$ will diminish as evident in Fig. 5b, but the exponential term will still dominate for σ_R less than or equal to $\langle R\rangle$. We note that the case for $\sigma_R > \langle R\rangle$ leads to unphysical nonzero values of R for the example of a normal distribution.

For the case of $k_d \to 0$, we note in Eq. (5) that the power term tends to 1, the exponential terms tend to $1 - k_d R$, and $1/\Gamma(k_d) \approx k_d$ to give the linear dependence in Eq. (9).

Exponentially distributed waiting times for transcription factor kinetics

We described the production of protein in bursts, which, in one example, may result from repressor binding fluctuations. We describe here when the assumption of exponential kinetics for repressor rebinding may be valid.

The kinetics of transcription factor binding and unbinding from specific sites on DNA involve free diffusion as well as interaction and diffusion along nonspecific DNA sequences. The full treatment of this problem requires consideration of spatial correlations on short time scales for both one- and three-dimensional diffusion.[43] Even if we ignore one-dimensional sliding along DNA and approximate the dynamics as binding and unbinding between freely diffusing particles, the remaining spatial correlation results in nonexponential dissociation and association kinetics. This is because the microscopic binding and unbinding by the same two particles can occur many times before a macroscopic dissociation and interchange of particles.

Assume that there is one DNA binding site on a promoter and a concentration of c repressor molecules per cell volume. If the promoter is initially bound by a repressor that then dissociates, the mean time before rebinding will be $T_1 = (ck_a(1 + k_a^m/4\pi rD))^{-1}$, assuming the repressor undergoes only free diffusion with diffusion constant D and that the binding site has a characteristic size of r.[29] Here, k_a^m is the microscopic association rate for repressors directly adjacent to the binding site. In the case of $k_a^m \gg 4\pi rD$, the rebinding time is clearly different from the macroscopic $T_2 = (ck_a)^{-1}$ because on short time scales, the same repressor that dissociated rapidly rebinds, while on long time scales, the repressors from the entire cell volume contribute to rebinding. If $T_1 \ll T_2$, we expect that both the mean and distribution of waiting times should vary significantly from monoexponential kinetics. In the case of $k_a^m \ll 4\pi rD$, though, rebinding is slow and not limited by diffusion. An example for this case would be if small molecule modulated the binding properties of the repressor. The repressor may only be able to recognize the specific binding site if the small molecule is released from the repressor. Then, $T_1 \approx T_2$ and we expect that the mean and distribution of waiting times should agree with the conventional exponential kinetics.

Another situation that can result in $k_a^m \ll 4\pi rD$ is competitive binding for the promoter region by both the repressor and RNA polymerase. Once bound to the promoter, the transition from initiation to elongation for an *E. coli* RNA polymerase may take 1–3 s.[44] During this time, occupancy of the promoter by RNAP would prevent rebinding by any repressor molecules, and even the most recently bound repressor molecule would lose its spatial correlation with the original binding site. The *in vivo* diffusion constant of a typical bacterial repressor is 0.4 $\mu m^2/s$,[28] so that several seconds would be sufficient for a repressor molecule to diffuse over a large macroscopic region compared to the *E. coli* dimension of ~1 μm. Thus, if even just one RNA polymerase displaces the repressor for binding to the promoter, use of the macroscopic $T_2 = (ck_a)^{-1}$ for rebinding of the next repressor molecule would be appropriate.

Calculation of p_{switch}

For the case of high repression, so that $K_{eq} \gg 1$, we can use Eq. (6) to calculate p_{switch}. In this case, $a \sim k_d$ as well; thus, we can substitute Eq. (6) into Eq. (1) and obtain:

$$p_{switch} = \frac{1}{a}(c/b)^a \frac{e^{-c/b}}{\Gamma(a)} \quad (21)$$

The values of a and b can be derived from transcription kinetics as described in Eq. (4). Experimentally, if it is possible to measure the switching rate, k, and the rate of any *in vivo* molecular reaction, k_{molec}, independently, then $p_{switch} = k/k_{molec}$ for that molecular reaction.

Use of experimental data

In our analysis of the *lac* operon switching, we made use of previously published data.[13] From this reference, the data of Fig. 1d were used to determine a_{small} and b_{small}. The value of the threshold c was taken directly from the text and Fig. 2b. The upper bound on a_{large} is taken from the number of bursts greater than 10 molecules in Fig. 3b. The rate of switching is calculated from Fig. 1c, with 10% switched after 24 h and a cell cycle of 1 h yielding a switching rate of 0.004 per cell cycle.

Acknowledgements

We would like to thank Long Cai for initial discussions motivating this work. This work is funded by the National Institutes of Health (E.I.S.). P.J.C. acknowledges the John and Fannie Hertz Foundation.

Appendix A. Feedback is necessary for extending burst size in the *lac* operon

We examined switching in the *lac* operon using a simplified description of small and large bursts. Although such an approach may provide valid qualitative and even quantitative predictions, we consider the significance of mechanistic details that may have been left out.

Burst size is a conceptual tool for the case of $p_{switch} \sim 1$, since cells with a large burst switch to a state of continual high expression, producing a burst of undefined size. However, the burst size remains a key parameter connecting transcription factor kinetics with switching. In an *E. coli* strain lacking permease feedback and, hence, unable to switch, a small number of large bursts were observed directly with a mean of 80 molecules.[13] This mean is smaller than the necessary burst sizes calculated in Fig. 6 to generate the correct switching rate for the wild-type strain, suggesting that the burst size may depend on the presence of permease feedback.

Because large bursts were observed to last for many minutes, in the presence of permease feedback, the intracellular inducer concentration could change significantly during the time of a single large burst. Then, a higher intracellular inducer concentration would further reduce the association rate of the repressor. Thus, although switching would be initiated by a dissociation event that is independent of permease feedback, both the inducer and permease have critical functions during the actual burst event itself. In other words, strains with or without permease feedback may have the same a_{large} but different b_{large}.

Thus, while our analysis of the *lac* operon provides some mechanistic insight into its switching behavior, there are even more molecular details to explore.

References

1. Ito, T., Chiba, T., Ozawa, R., Yoshido, M., Hattori, M. & Sakaki, Y. (2001). A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl Acad. Sci. USA*, **98**, 4569–4574.
2. Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D. & Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science*, **298**, 824–827.
3. Butland, G., Peregrín-Alvarez, J. M., Li, J., Yang, W., Yang, X., Canadien, V. *et al.* (2005). Interaction network containing conserved and essential protein complexes in *Escherichia coli*. *Nature*, **433**, 531–537.
4. Novick, A. & Weiner, M. (1957). Enzyme induction as an all-or-none phenomenon. *Proc. Natl Acad. Sci. USA*, **43**, 553–566.
5. Eisenstein, B. I. (1981). Phase variation of type 1 fimbriae in *Escherichia coli* is under transcriptional control. *Science*, **214**, 337–339.
6. Nester, E. W. & Stocker, B. A. D. (1963). Biosynthetic latency in early stages of deoxyribonucleic acid transformation in *Bacillus subtilis*. *J. Bacteriol.* **86**, 785–796.
7. Artavanis-Tsakonas, S., Rand, M. D. & Lake, R. J. (1999). Notch signaling: cell fate control and signal integration in development. *Science*, **284**, 770–776.
8. Watt, F. M. & Hogan, B. L. M. (2000). Out of Eden: stem cells and their niches. *Science*, **287**, 1427–1430.
9. Gardner, T., Cantor, C. R. & Collins, J. J. (2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, **403**, 339–342.
10. Kauffman, S., Peterson, C., Samuelsson, B. & Troein, C. (2003). Genetic networks with canalyzing Boolean rules are always stable. *Proc. Natl Acad. Sci. USA*, **101**, 17102–17107.
11. Mettetal, J. T., Muzzey, D., Pedraza, J. M., Ozbudak, E. M. & van Oudenaarden, A. (2006). Predicting stochastic gene expression dynamics in single cells. *Proc. Natl Acad. Sci. USA*, **103**, 7304–7309.
12. Maamar, H., Raj, A. & Dubnau, D. (2007). Noise in gene expression determines cell fate in *Bacillus subtilis*. *Science*, **317**, 526–529.
13. Choi, P. J., Cai, L., Frieda, K. & Xie, X. S. (2008). A stochastic single-molecule event triggers phenotype switching of a bacterial cell. *Science*, **322**, 442–446.
14. Rao, C. V., Wolf, D. M. & Arkin, A. P. (2002). Control, exploitation, and tolerance of intracellular noise. *Nature*, **420**, 231–237.
15. Raser, J. M. & O' Shea, E. K. (2005). Noise in gene expression: origins, consequences, and control. *Science*, **309**, 2010–2013.
16. Dubnau, D. & Losick, R. (2006). Bistability in bacteria. *Mol. Microbiol.* **61**, 562–572.
17. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. (2002). Stochastic gene expression in a single cell. *Science*, **297**, 1183–1186.
18. Raser, J. M. & O' Shea, E. K. (2004). Control of stochasticity in eukaryotic gene expression. *Science*, **304**, 1811–1814.
19. Golding, I., Paulsson, J., Zawilski, S. M. & Cox, E. C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell*, **123**, 1025–1036.
20. Blake, W. J., KÆrn, M., Cantor, C. R. & Collins, J. J. (2003). Noise in eukaryotic gene expression. *Nature*, **422**, 633–637.
21. Walczak, A. M., Onuchic, J. N. & Wolynes, P. G. (2005). Absolute rate theories of epigenetic stability. *Proc. Natl Acad. Sci. USA*, **102**, 18926–18931.
22. Morelli, M. J., Allen, R. J., Tanase-Nicola, S. & ten Wolde, P. R. (2008). Eliminating fast reactions in

stochastic simulations of biochemical networks: a bistable genetic switch. *J. Chem. Phys.* **128**, 045105.
23. Aurrell, E. & Sneppen, K. (2002). Epigenetics as a first exit problem. *Phys. Rev. Lett.* **88**, 048101.
24. Mehta, P., Mukhopadphyay, R. & Wingreen, N. S. (2008). Exponential sensitivity of noise-driven switching in genetic networks. *Phys. Biol.* **5**, 026005.
25. Little, J. W., Shepley, D. P. & Wert, D. W. (1999). Robustness of a gene regulatory circuit. *EMBO J.* **18**, 4299–4307.
26. Mandelstam, J. (1958). Turnover of protein in growing and non-growing populations of *Escherichia coli*. *Biochem. J.* **69**, 110–119.
27. Friedman, N., Cai, L. & Xie, X. S. (2006). Linking stochastic dynamics to population distributions: an analytical framework of gene expression. *Phys. Rev. Lett.* **97**, 168302.
28. Elf, J. & Li, G. W. (2007). Probing transcription factor dynamics at the single-molecule level in a living cell. *Science*, **316**, 1191–1194.
29. Berg, O. G. (1978). On diffusion-controlled dissociation. *Chem. Phys.* **31**, 47–57.
30. Kepler, T. B. & Elston, T. C. (2001). Stochasticity in transcription regulation: origins, consequences, and mathematical representations. *Biophys. J.* **81**, 3116–3136.
31. Paulsson, J. (2005). Models of stochastic gene expression. *Phys. Life Rev.* **2**, 157–175.
32. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. (2005). Gene regulation at the single cell level. *Science*, **307**, 1962–1965.
33. Newman, J. R., Ghaemmaghami, S., Ihmels, J., Breslow, D. K., Noble, M., DeRisi, J. L. & Weissman, J. S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature*, **441**, 840–846.
34. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & van Oudenaarden, A. (2004). Multistability in the lactose utilization network of *Escherichia coli*. *Nature*, **427**, 737–740.
35. Yu, J., Xiao, J., Ren, X., Lao, K. & Xie, X. S. (2006). Probing gene expression in live cells, one protein molecule at a time. *Science*, **311**, 1600–1603.
36. Cai, L, Friedman, N. & Xie, X. S. (2006). Stochastic protein expression in individual cells at the single molecule level. *Nature*, **440**, 358–362.
37. Kussell, E. & Leibler, S. (2005). Phenotypic diversity, population growth, and information in fluctuating environments. *Science*, **309**, 2075–2078.
38. Wolf, D. W., Vazirani, V. V. & Arkin, A. P. (2005). Diversity in times of adversity: probabilistic strategies in microbial survival games. *J. Theor. Biol.* **234**, 227–253.
39. Dekel, E. & Alon, U. (2005). Optimality and evolutionary tuning of the expression level of a protein. *Nature*, **436**, 588–592.
40. Shahrezaei, V. & Swain, P. (2008). Analytical distributions for stochastic gene expression. *Proc. Natl Acad. Sci. USA*, **105**, 17251–17256.
41. Sanchez, A. & Kondev, J. (2008). Transcriptional control of noise in gene expression. *Proc. Natl Acad. Sci. USA*, **105**, 5081–5086.
42. Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361.
43. Berg, O. G., Winter, R. B. & von Hippel, P. H. (1981). Diffusion-driven mechanisms of protein translocation on nucleic acids. 1. Models and theory. *Biochemistry*, **21**, 6929–6948.
44. Wade, J. T. & Struhl, K. (2008). The transition from transcriptional initiation to elongation. *Curr. Opin. Genet. Dev.* **18**, 130–136.