Predicting stochastic gene expression dynamics in single cells

Jerome T. Mettetal*†, Dale Muzzey*†‡, Juan M. Pedraza*, Ertugrul M. Ozbudak*, and Alexander van Oudenaarden*§

*Department of Physics, Massachusetts Institute of Technology, Cambridge, MA 02139; and ‡Harvard University Graduate Biophysics Program, Harvard Medical School, Boston, MA 02115

Edited by Nancy J. Kopell, Boston University, Boston, MA, and approved March 19, 2006 (received for review November 14, 2005)

Fluctuations in protein numbers (noise) due to inherent stochastic effects in single cells can have large effects on the dynamic behavior of gene regulatory networks. Although deterministic models can predict the average network behavior, they fail to incorporate the stochasticity characteristic of gene expression, thereby limiting their relevance when single cell behaviors deviate from the population average. Recently, stochastic models have been used to predict distributions of steady-state protein levels within a population but not to predict the dynamic, presteady-state distributions. In the present work, we experimentally examine a system whose dynamics are heavily influenced by stochastic effects. We measure population distributions of protein numbers as a function of time in the *Escherichia coli* lactose uptake network (*lac* operon). We then introduce a dynamic stochastic model and show that prediction of dynamic distributions requires only a few noise parameters in addition to the rates that characterize a deterministic model. Whereas the deterministic model cannot fully capture the observed behavior, our stochastic model correctly predicts the experimental dynamics without any fit parameters. Our results provide a proof of principle for the possibility of faithfully predicting dynamic population distributions from deterministic models supplemented by a stochastic component that captures the major noise sources.

gene networks | systems biology

One of the central goals of systems biology is to predict the dynamic behavior of a cell's genetic and metabolic networks. These predictions traditionally stem from models in which discrete molecular events, such as transcription and translation, are represented by continuous and deterministic differential equations. Such equations are valid when behavior of individual cells is very similar to the average behavior of the population. However, in many cases, the inherently stochastic nature of biological systems leads to significant cell-to-cell variability (1–3), and previous studies indicate that individual cells often behave very differently from the population average in response to external stimuli. For example, studies of bacterial persistence indicate that the population survival rate can be fundamentally different from the average cell's survival rate in response to environmental stress (4, 5). The impact of noise-induced population heterogeneity is also relevant when studying cellular memory, where fluctuations in protein numbers can cause the stability of epigenetic memory to degrade by effectively causing cells to forget their original states (6, 7). In these cases, stochastic modeling techniques must be used to describe the large cell-to-cell variability.

Although deterministic models can describe dynamic network behavior and analytical stochastic models can faithfully predict steady-state population distributions (8), little work has been done to connect dynamic cellular behavior with noise models. It is important that stochastic models of biological systems correctly capture system dynamics because few biological systems ever reach steady state. Prior pioneering studies aimed at predicting stochastic dynamics have modeled stochastic effects by including high levels of microscopic detail (9). Although this approach is correct in principle, it is often too complicated to have general applicability because the parameters required are usually unmeasured or difficult to acquire. Since the advent of these microscopic approaches, much has been learned about sources and propagation of noise in gene networks (8, 10–21), leading to comprehensive models of stochastic behavior. However, these previous models have lacked a sufficiently detailed set of dynamic data on which to test predictions of dynamic population distributions (22–24). Thus, it has still not been demonstrated that the current understanding of noise, which accurately describes distributions of protein concentrations in steady state, can be applied to predict dynamic distributions reflecting noise-induced behavior.

In this work, we investigate the predictive power of stochastic dynamics by using an integrated experimental and computational approach in which we construct a stochastic model of cellular dynamics. To test the model's predictions, we experimentally measure population distributions of protein levels over time in the lactose uptake network of *Escherichia coli* and then compare these data to the predicted population distributions. To construct our predictive model, we first build a deterministic model that incorporates relevant network components. Next, we use steady state measurements of this network to characterize the magnitude of the relevant noise sources. Finally, by combining these noise sources with the deterministic model, we create a dynamic stochastic model that is able to predict the dynamic behavior of distributions. Using this technique, we show that once macroscopic rates are known, we only need two additional parameters that characterize the noise in each gene to faithfully predict experimental dynamic population distributions without any fit parameters.

The Lactose Uptake Network

At the systems level, the lactose uptake network in single *E. coli* cells displays an "all or nothing" response depending on the extracellular concentration of the inducer (25, 26). This ability to display different phenotypes at a single extracellular inducer concentration has been attributed to a positive feedback loop, which is shown in Fig. 1*a*. LacY (indicated by purple in Fig. 1) is a transmembrane protein involved in active uptake of the inducer thiomethylgalactoside (TMG) (indicated by orange in Fig. 1). The synthesis of LacY is under the control of the *lac* promoter, which is repressed by the lac repressor, LacI (indicated by blue in Fig. 1), in the absence of TMG. However, intracellular TMG molecules bind to LacI tetramers, causing their dissociation from DNA and an increase in *lacY* transcrip-

Conflict of interest statement: No conflicts declared.

This paper was submitted directly (Track II) to the PNAS office.

Abbreviations: TMG, thiomethylgalactoside; CRP, cAMP receptor protein; RFP, red fluorescent protein.

†J.T.M. and D.M. contributed equally to this work.

§To whom correspondence should be addressed at: Department of Physics, Room 13-2008, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139. E-mail: avano@mit.edu.

© 2006 by The National Academy of Sciences of the USA

www.pnas.org/cgi/doi/10.1073/pnas.0509874103

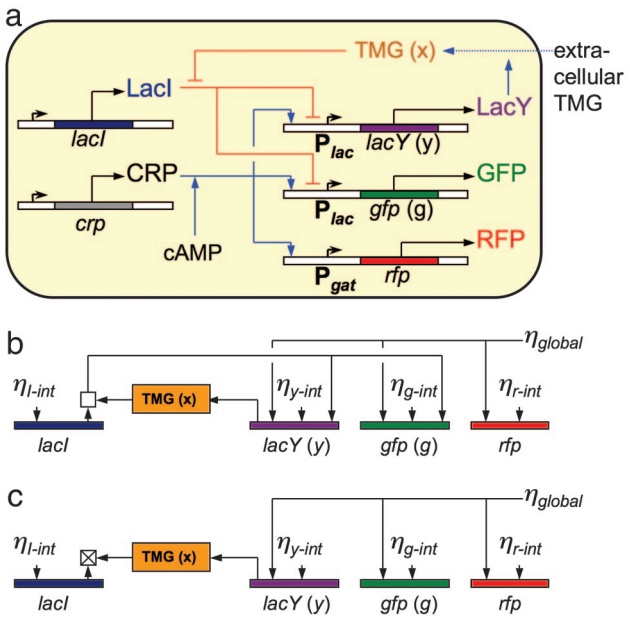


Fig. 1. Network and noise diagram. (*a*) Diagram of the lactose utilization network. Blue arrows indicate positive interactions, red bars indicate negative interactions, and black arrows denote protein production. A positive feedback loop from LacY to TMG to LacI back to LacY creates the potential for multistability (high and low steady states). The fluorescent reporter GFP integrated in the genome is expressed in parallel with LacY under control of the lac promoter and reports the induction level of the cell. RFP under control of the *gat* promoter reports activity of the activator CRP. (*b*) The noise network for the lactose utilization network. Intrinsic noise is fed into each protein level and is propagated through the network. The square above LacI represents the combination and propagation of noise from total LacI and TMG through the active fraction of LacI tetramers, which depends on the concentration of intracellular TMG. (*c*) The effective noise network for induced cells with high levels of intracellular TMG, where LacI tetramers are highly inactivated. The crossed square represents the effective inactivation of this feedback by increased levels of intracellular TMG.

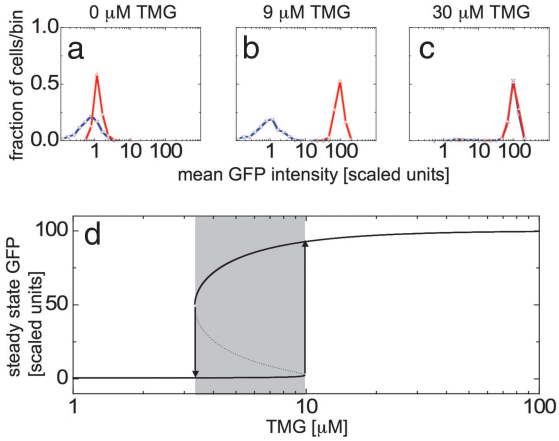


Fig. 2. Demonstration of hysteresis in the long time limit. Histograms of mean GFP fluorescence are shown for cells with a fully induced (red) and fully uninduced (blue) history resuspended and grown for 20 h in 0 (*a*), 9 (*b*), and 30 (*c*) μM TMG. Induced cells grown in 0 μM TMG for 20 h still contain slightly higher quantities of GFP than uninduced cells, and this difference is roughly equivalent to that expected from exponential decay of fluorescence due to dilution of GFP during cell division. (*d*) Steady-state solutions of the deterministic model. The induced state is shown as the upper dark line whereas the uninduced state is shown as the lower dark line. The intermediate unstable steady state is shown as a dashed line in the shaded bistable region. Cells remain in either the induced or uninduced states until they are moved to a concentration of inducer at which the previous state is unstable (vertical arrows).

tion. *lacY* transcription can be further activated by the cAMP receptor protein (CRP) (indicated by gray in Fig. 1), which upon association with cAMP binds to an activator site in the *lac* promoter and increases the probability of transcription. In summary, this positive feedback loop is composed of two negative connections and one positive connection: TMG inhibits LacI, LacI represses *lacY* transcription, and LacY increases the intracellular TMG concentration. In addition to this natural endogenous network, we constructed two fluorescent reporter systems to monitor the state of the network in single cells. The gene encoding for GFP (*gfp*) (indicated by green in Fig. 1) under control of the *lac* promoter was integrated into the genome and reports the concentration of LacY. Additionally, we placed a red fluorescent protein (RFP) gene (*rfp*) (indicated by red in Fig. 1) under the control of the *gat* promoter on a plasmid, which contains a CRP activation region and a gat repressor binding site. Because wild-type K12 *E. coli* strains lack a functional gat repressor protein, GatR (27), RFP is a faithful reporter for the activity of CRP (28).

Experimental Results

To model the stochastic dynamics of the lactose uptake network, it is first necessary to understand single-cell behavior. At steady state, the positive feedback loop causes cells to be in either an ON (induced) state, during which they maximally produce LacY and GFP, or an OFF (uninduced) state, during which LacY and GFP are produced at a minimal, basal rate. In ON cells, LacY imports enough extracellular TMG to maximally produce LacY, whereas OFF cells do not have enough LacY or extracellular TMG to produce LacY molecules faster than they are lost. These circumstances mean that cells do not generally contain intermediate concentrations of LacY and GFP when in steady state.

To observe this behavior, we prepare cells in either the ON or OFF state by growing them for 24 h in media with 100 μM TMG or 0 μM TMG, respectively. We then remove the cells from this preparation media and subsequently grow them in fresh media containing an intermediate concentration of extracellular TMG. Cells are grown for 20 h (approximately seven cell generations), at which point the population distributions are no longer changing quickly. For cultures resuspended in very high or low concentrations of TMG, cells occupy either the ON or OFF state, respectively, independent of their induction history (Fig. 2 *a* and *c*). However, resuspension in intermediate concentrations of TMG maintain an ON or OFF population for extended periods of time in either the higher or lower peak, respectively (Fig. 2*b*), indicative of hysteresis (Fig. 2*d*).

Although we find that individual cells are either in the OFF or ON states after 20 h, measurements at shorter time intervals must reveal transient, intermediate distributions that reflect the population morphing from its initial state (Fig. 2*a*, blue curve) to its final state (Fig. 2*c*, blue curve). To characterize the dynamics of these population distributions in response to changes in inducer concentration, we sample the population at various times and measure GFP levels in single cells. Fully induced or uninduced cells are washed and subsequently resuspended in media with an intermediate concentration of TMG. Next, the mean GFP fluorescence levels of individual cells are measured. Histograms are generated every 1 or 2 hours for several hours after resuspension in media with intermediate TMG concentration.

Two types of dynamic responses are observed: ballistic and stochastic. An example of a ballistic transition is shown in Fig. 3*a*, where ON cells resuspended in 0 μM TMG collectively switch OFF, drifting toward the new stable state. By contrast, Fig. 3*b*

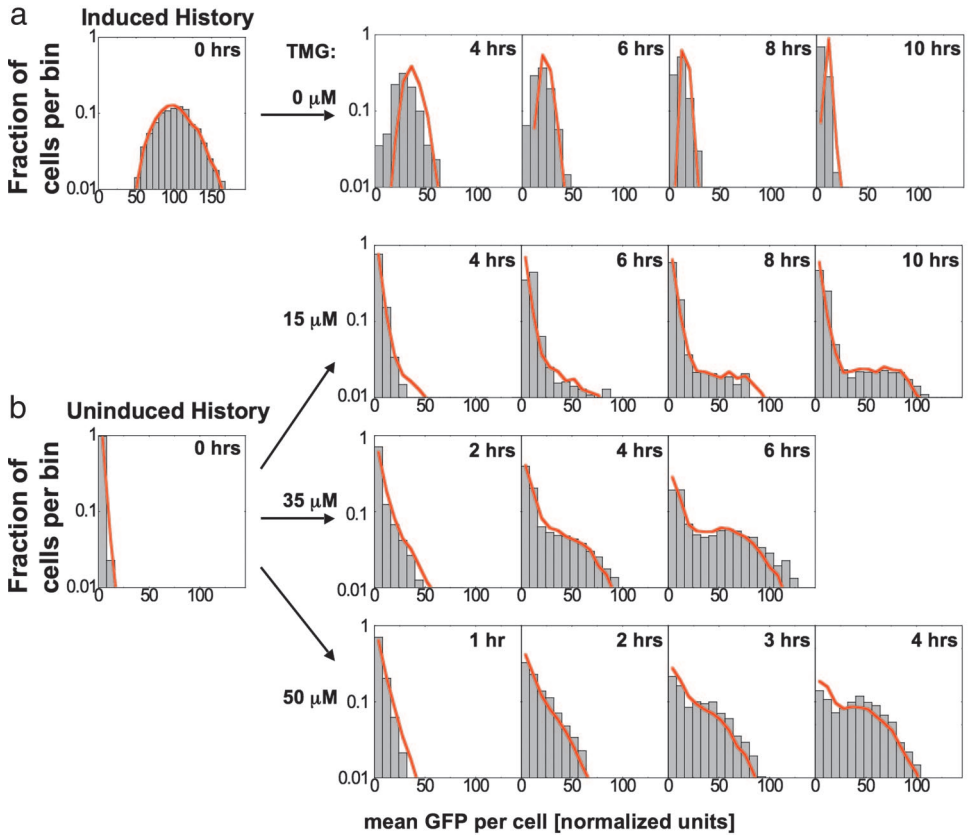


Fig. 3. Comparison between our stochastic model and experiments for transitions between OFF and ON steady states. Gray boxes are histograms of single cell GFP fluorescence for populations of 2,000–6,000 cells, and red lines represent a population of 10,000 simulated cells using a Monte-Carlo algorithm. (*a*) ON cells grown in 50 μM TMG and then placed in 0 μM TMG transition as a uniform population to the uninduced state with a single ''ballistic'' peak. (*b*) OFF cells grown at 0 μM TMG are transferred to 15, 35, and 50 μM TMG. These three populations display stochastic switching behavior, where cells randomly leave the uninduced state and move toward the induced state. Simulations predict the ballistic behavior associated with cells turning OFF and the stochastic behavior associated with cells turning ON. In both cases, the experimental distributions are well matched by the model without any additional fit parameters.

illustrates stochastic transitions, where some initially OFF cells remain OFF while a subpopulation switches to the ON state. Characteristic of a stochastic response, the OFF peak decreases in magnitude exponentially with time.

Deterministic Model

Several deterministic models have been used to explain the bistable behavior observed in the lactose utilization network (28–31). We augment a model that has been used to describe the strains analyzed in this study (28). Our model is composed of three equations:

$$\tau_x \frac{dx}{dt} = \beta(TMG)y + \lambda \cdot TMG - x, \qquad [1]$$

$$\tau_y \frac{dy}{dt} = \alpha \frac{1 + x^2}{\rho + x^2} - y, \qquad [2]$$

and

$$\tau_g \frac{dg}{dt} = \alpha \frac{1 + x^2}{\rho + x^2} - g. \qquad [3]$$

Here, x, y, and g represent intracellular concentrations of TMG, LacY, and GFP, respectively. *TMG* denotes the extracellular concentration of TMG. $1/\tau_x$ represents the rate of loss of intracellular TMG due to export, degradation, or dilution from cell division and growth. $1/\tau_y$ and $1/\tau_g$ represent the combined loss from dilution and degradation of LacY and GFP, respectively. The rate constant of active uptake of TMG per LacY molecule is proportional to β, which is a function of extracellular TMG, whereas λ represents passive uptake of TMG independent of LacY. We assume that GFP is transcribed at a rate identical to that of LacY because both are expressed under control of the *lac* promoter. Therefore, we set α as the maximal production rate of both LacY and GFP. ρ is the repression factor representing the ratio of transcription rates at the *lac* promoter between induced and uninduced cells. This factor accounts for the effect of fully activating all present LacI tetramers in the absence of intracellular TMG. Derivations of these equations can be found in the supporting information, which is published on the PNAS web site.

By using these equations, it has been shown that the system can have either a single stable steady state (monostable) or two stable steady states separated by an unstable steady state (bistable), depending on the concentration of extracellular TMG (Fig. 2*d*). Previously, the parameters α, β, and ρ have been determined by fitting the theoretical monostable–bistable boundaries (vertical arrows in Fig. 2*d*) to those measured experimentally (28). Because the network used in our study is identical, we will use the parameters α, β, and ρ as determined in this previous study (Table 1). We set τ_y and τ_g equal to the dilution time scale due to cell growth, $\tau_{1/2} = \tau_{\text{division}}$, because we assume that the active degradation rate of GFP and LacY is much smaller than the dilution rate due to cell growth; here, τ_{division} is the estimated average time between cell divisions. This

Table 1. Parameters used in the dynamic stochastic model

Parameter	Value	Error	Source
α	100	16	Ref. 28
β	0.123(TMG)$^{0.6}$	15%	Ref. 28
ρ	170	34	Ref. 28
λ	0.06	(0.03, 0.12)	This study
τ_x	0 min	35	This study
$\tau_{1/2}$	216 min	43	This study
N_{LacY}	790 proteins	210	This study
N_{GFP}	790 proteins	210	This study
N_{LacI}	50 proteins	—	Ref. 34
b_{LacY}	34.8 proteins	10.1	This study
b_{GFP}	34.8 proteins	10.1	This study
b_{LacI}	5 proteins	—	Ref. 34

—, not applicable.

approximation leaves τ_x and λ as the only undetermined parameters in the deterministic model.

We estimate the decay time of intracellular TMG, τ_x, by measuring how quickly the average GFP levels of a fully induced population decrease when placed in media without TMG (Fig. 3*a*). In this scenario, LacY no longer affects the dynamics of the cells because there is no extracellular TMG to complete the feedback loop. We assume the parameter τ_x to be much smaller than $\tau_{1/2}$ because cells cease GFP production within 10–20 min after removal of extracellular TMG (see the supporting information for further details), and a fit of the deterministic model to the decreasing GFP levels yields $\tau_x = 0$ min. This effect could be due to rapid loss of intracellular TMG through efflux, which is known to occur for other inducers (32, 33). Thus, we equilibrate Eq. **1** in relation to Eqs. **2** and **3** by setting: $\tau_x(dx/dt) = 0$.

In induced cells with many TMG-transporting LacY molecules, passive TMG uptake, λ, should contribute only a small fraction to the intracellular TMG concentration. However, in uninduced cells with few LacY molecules, this passive "leak" rate may be significant. To estimate this rate, the full deterministic model is fit to the population average of fully uninduced cells placed in 50 μM TMG, giving $\lambda = 0.06$ (see the supporting information).

Noise and Stochastic Measurements

To make predictions about switching transitions and dynamic population distributions, we must include the effect of stochastic fluctuations in our model. It has been shown (8, 12, 13, 15, 17, 18, 20, 21) that noise in protein levels comes mainly from discreteness of mRNA and protein numbers (intrinsic noise) as well as from global changes in intracellular environment that affect decay and production rates (global noise). To model intrinsic noise, we must estimate two parameters for each gene: the average number of proteins produced from a single mRNA (burst size) and the conversion factor between absolute protein numbers and fluorescence counts.

Fig. 1*b* contains a diagram indicating the generation and propagation of noise in the network. First, LacY, LacI, and GFP are sensitive to their corresponding intrinsic noise terms $\eta_{y\text{-int}}$, $\eta_{i\text{-int}}$, and $\eta_{g\text{-int}}$, respectively (Fig. 1*b*). These terms are due to the random creation and destruction events of *lacY*, *lacI*, and *gfp* mRNA and the corresponding proteins. The RFP reporter generates an intrinsic noise term, $\eta_{r\text{-int}}$ (Fig. 1*b*), which contains fluctuations due to RFP mRNA and RFP proteins as well as fluctuations in plasmid number. Noise generated by CRP and other factors, such as RNA polymerase and ribosomes, is combined into the term η_{global}, which we treat as a multiplicative factor on the production rates of LacY, GFP, and RFP.

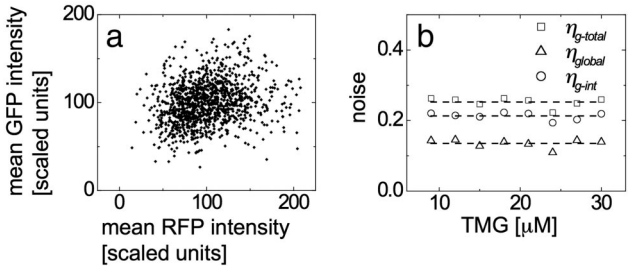


Fig. 4. Measurement of intrinsic and extrinsic noise is accomplished by comparing RFP and GFP concentrations in individual cells. (*a*) RFP and GFP concentrations for individual cells grown in media containing 30 μM TMG for 24 h. A slight correlation between the two concentrations is present, indicating that the two share a weak source of global noise. (*b*) Measurement of this correlation allows division of total noise in GFP into intrinsic and extrinsic components, each of which are calculated from 9 to 30 μM TMG.

Fluctuations in LacI numbers are propagated directly into LacY and GFP, and the strength of this transmission depends greatly on the intracellular TMG concentration because TMG binding decouples LacI from the production of LacY and GFP (8, 13). Finally, noise in LacY will cause fluctuations in intracellular TMG concentration that affect the binding of LacI to the *lac* promoter. This effect causes LacY fluctuations to be transmitted into both LacY and GFP noise with a magnitude dependent on TMG concentration.

Extrinsic noise can be determined by examining correlations between levels of proteins influenced by the same upstream regulators. To extract the noise parameter, η_{global}, from the distribution of GFP concentrations, it would generally be necessary to solve a set of equations describing the propagation of noise through the entire network (8). However, the network can be simplified greatly by considering only fully induced cells, where TMG-bound and inactivated LacI no longer affects GFP and LacY production. In this case, fluctuations in LacY, GFP, and RFP expression levels are dependent only on extrinsic noise levels, η_{global}, as well as each protein's intrinsic noise level $\eta_{y\text{-int}}$, $\eta_{g\text{-int}}$, and $\eta_{r\text{-int}}$ (Fig. 1*c*). Because the term η_{global} is shared by GFP and RFP, it is possible to separate the total GFP noise into intrinsic and extrinsic components (see the supporting information for details). GFP and RFP fluorescence levels are measured in individual cells on several populations induced with different concentrations of extracellular TMG. For 30 μM TMG, this distribution is shown in Fig. 4*a*, indicating a weak correlation between GFP and RFP levels. For GFP, the total noise, $\eta_{g\text{-total}}$, and extrinsic noise, η_{global}, are measured directly from the GFP and RFP distributions in the induced population using the relations

$$\eta^2_{g-\mathrm{total}} = \frac{\langle \delta G^2 \rangle}{\langle G \rangle^2} = \frac{\langle G^2 \rangle - \langle G \rangle^2}{\langle G \rangle^2} \qquad [4]$$

and

$$\eta^2_{\mathrm{global}} = \frac{\langle \delta G \, \delta R \rangle}{\langle G \rangle \langle R \rangle} = \frac{\langle GR \rangle - \langle G \rangle \langle R \rangle}{\langle G \rangle \langle R \rangle}. \qquad [5]$$

Here brackets $\langle \ldots \rangle$ represent the population average of fluorescent levels of single cells in the ON state only. The intrinsic GFP noise, $\eta^2_{g\text{-int}} = \eta^2_{g\text{-total}} - \eta^2_{\mathrm{global}}$, is calculated for extracellular TMG ranging from 9 to 30 μM (Fig. 4*b*), and mean values of the noise measurements are $\eta_{g\text{-total}} = (0.25 \pm 0.04)$, $\eta_{g\text{-int}} = (0.21 \pm 0.03)$, and $\eta_{\mathrm{global}} = (0.14 \pm 0.02)$. These values remain constant over the range of measurement, indicating that the ON cells have similar noise characteristics regardless of the concentration of extracellular TMG.

BIOPHYSICS

To calibrate between fluorescence and the absolute number of molecules, it is sufficient to determine the number of GFP molecules in an induced cell, N. Rosenfeld *et al.* (18) showed that intensity fluctuations introduced by cell division vary with the number of fluorescent molecules in the cell. Upon cell division, each molecule will be independently and randomly partitioned into one of the two daughter cells. This process can be described by a binomial distribution, where the difference between the numbers of molecules in each daughter cell will, on average, be proportional to $N^{1/2}$, meaning that as the number of molecules decreases (increases) the fractional asymmetry introduced by division, $N^{1/2}/N = N^{-1/2}$, will increase (decrease). Therefore, by using this as a calibration tool, we measure the mean GFP fluorescence in pairs of cells that have recently undergone cell division and estimate that the average number of GFP proteins in an induced cell is $N_{\mathrm{GFP}} = 790 \pm 210$ molecules (see the supporting information for details).

We next estimate the second missing noise parameter, burst size, which can be determined from the intrinsic noise, $\eta_{g\text{-int}}$, by using the relation $\langle \delta n^2 \rangle = \langle n \rangle (b + 1)$ (11, 12), where n is the number of protein molecules and b is the average number of proteins produced from a single mRNA. This relation stems from the fact that translation effectively amplifies the noise associated with low levels of mRNA. We find that the burst size for GFP is $b_{\mathrm{GFP}} = 35.3 \pm 9.7$, consistent with other burst size measurements in *E. coli* (8). Because the same promoter is regulating GFP and LacY expression, we assume that the production rate of mRNA should be similar for the two proteins. Thus, we set $(N_{\mathrm{LacY}}/b_{\mathrm{LacY}}) = (N_{\mathrm{GFP}}/b_{\mathrm{GFP}})$, where N_{LacY} is the number of LacY molecules in a fully induced cell, and b_{LacY} is the burst size of a LacY mRNA. We could reduce the ratio $N_{\mathrm{LacY}}/b_{\mathrm{LacY}}$ to a single parameter analytically because the mRNA production rate for LacY is proportional to $N_{\mathrm{LacY}}/b_{\mathrm{LacY}}$, and the burst size in units of fluorescence is proportional to $(N_{\mathrm{LacY}}/b_{\mathrm{LacY}})^{-1}$. However, to proceed later with simulations that model explicit molecular events, we need to assign values for both N_{LacY} and b_{LacY}. Therefore, without loss of generality, we choose values for these parameters that maintain the required ratio by setting them equal to those measured for GFP: $N_{\mathrm{LacY}} = N_{\mathrm{GFP}}$ and $b_{\mathrm{LacY}} = b_{\mathrm{GFP}}$.

We expect noise to be transmitted from the LacI component as well, but to analyze these contributions it is necessary to examine steady state distributions around the uninduced fixed point. However, a small signal to noise ratio combined with increasing nonlinearity at these low concentrations makes extraction of relevant information difficult. Instead, we use published estimates of the molecule counts and burst sizes of LacI, which we interpret as $b_{\mathrm{LacI}} = 5$ and $N_{\mathrm{LacI}} = 50$ in our strain (34).

Stochastic Model and Predictions

We now have a full set of parameters for the dynamic deterministic model, as well as all parameters necessary to describe the steady state noise properties of the lactose uptake network. To determine whether these parameters can be used to predict the full dynamic distributions measured in the experimental section, we construct a dynamic stochastic model. In principle, we have already fully defined the model, but for simulation purposes we use a reduced model that contains only essential events. The model consists of three processes: (*i*) mRNA production followed immediately by a burst of protein production and mRNA degradation, (*ii*) protein degradation, and (*iii*) extrinsic or global noise. We therefore ignore specific events, such as binding of LacI or CRP to operator sites and timing between productions of individual protein molecules. These processes and the corresponding rates are summarized in Table 2 and described in greater detail in the supporting information. We use a modified Gillespie's Monte-Carlo method (35) to simulate large populations of individual cells. Fig. 3*a* shows the results of the simulation (red solid lines) for induced-to-uninduced transitions where induced cells are placed in 0 μM TMG. Fig. 3*b* shows simulations (red solid lines) of uninduced cells grown in 15, 35, and 50 μM TMG, respectively.

Table 2. Stochastic Model

Protein	Event	Reaction	Rate
GFP	Burst	$G \rightarrow G + B(b_{\mathrm{GFP}})$	$f(x(Y), I)\, N_{\mathrm{GFP}}/(b_{\mathrm{GFP}}\, \tau_{1/2})\varepsilon$
	Decay	$G \rightarrow G - 1$	$G/\tau_{1/2}$
LacY	Burst	$Y \rightarrow Y + B(b_{\mathrm{LacY}})$	$f(x(Y), I)\, N_{\mathrm{LacY}}/(b_{\mathrm{LacY}}\, \tau_{1/2})\, \varepsilon$
	Decay	$Y \rightarrow Y - 1$	$Y/\tau_{1/2}$
LacI	Burst	$I \rightarrow I + B(b_{\mathrm{LacI}})$	$N_{\mathrm{LacY}}/(b_{\mathrm{LacY}}\, \tau_{1/2})$
	Decay	$I \rightarrow I - 1$	$I/\tau_{1/2}$

$f(x, I) = (\{[(\rho - 1)I]/[N_{\mathrm{LacI}}(1 + x^2)]\} + 1)^{-1}$ with $x(Y) = Y\,(\alpha\beta/N_{\mathrm{LacY}}) + \lambda \cdot TMG$ models the effect of LacI repression based on instantaneous values of Y and I. ε represents a correlated extrinsic noise, with $\langle \varepsilon \rangle = 1$ and $\langle \varepsilon(t)\varepsilon(t + \Delta t) \rangle = 2\eta^2_{\mathrm{global}} \exp(-\Delta t/\tau_{1/2})$. Capital variables are absolute molecule numbers of the respective lowercase concentrations scaled to the deterministic model as defined in the supporting information: $G \equiv g\, N_{\mathrm{GFP}}/\alpha$ and $Y \equiv y\, N_{\mathrm{GFP}}/\alpha$. $B(b)$ is an exponentially distributed random integer with mean equal to burst size b.

The stochastic model reproduces ballistic and stochastic switching, whereas the deterministic model does not distinguish between these types of behavior. For the case shown in Fig. 3*a*, the experiments and simulations both show that every cell moves like the average obtained from the deterministic model. In addition to demonstrating average behavior, the stochastic model also correctly predicts the widths of the distributions. For the data shown in Fig. 3*b*, the individual cells behave very differently from the average, and the stochastic simulation captures this behavior. In this case, the peak at uninduced GFP levels slowly decays as a subpopulation of cells begins to transition to the induced state. In addition to demonstrating the general behavior, the model matches the rate of transitions out of the uninduced peak and predicts the shape of the population distribution over a wide range of TMG and time.

It is noteworthy that ballistic switching does not always occur when the initial state becomes unstable. For instance, although the OFF state is no longer stable in media with 50 μM TMG, the timing of an OFF-to-ON switching event depends on a cell's rare production of LacY mRNA and subsequent protein creation. It is this rare, stochastic burst of LacY that ultimately triggers the positive feedback loop and drives the OFF-to-ON transition. Conversely, the ON-to-OFF transition is ballistic because it requires the dilution of intracellular TMG and GFP, both of which are low-noise events involving high molecule numbers.

Discussion

We introduced a method that should have general applicability for the prediction of stochastic cellular dynamics. The first step includes characterization of a deterministic model that matches known steady-state behavior. Rate constants in this model can be estimated from published values, by fitting to steady state measurements, or through direct biochemical assays. Next, the magnitudes of noise sources are extracted from distributions of fluorescent counts and correlations between different expression reporters measured in steady state. The sources of intrinsic noise are then characterized by the discrete molecule numbers, N, and mRNA burst sizes, b. By combining these reaction rates and noise sources, a stochastic model is produced containing three important stochastic factors: mRNA production, protein degradation, and global noise.

The model is in good agreement with experimental data and can predict the type of response (ballistic versus stochastic), the escape rates from a state, and the distribution of reporter fluorescence values without any fit parameters. Furthermore, the predicted distributions proved to be robust against param-

eter variation in both magnitude and general behavior (see the supporting information). The precision of our model's predictions could be improved by careful measurement of individual model parameters; however, we have shown that previously obtained parameter estimates are sufficient to provide interesting quantitative information about network behavior not available from deterministic equations alone.

Materials and Methods

We calculate the experimental error in noise measurements [$\eta_{g\text{-total}} = (0.25 \pm 0.04)$, $\eta_{g\text{-int}} = (0.21 \pm 0.03)$, $\eta_{global} = (0.14 \pm 0.02)$] by setting error bars equal to the standard deviation of the noise values measured during independent experiments from 9 to 30 μM TMG. We determine the error in the value of b_{GFP} by calculating propagated errors from N_{GFP} and $\eta_{g\text{-int}}$.

Model predictions were generated by Monte-Carlo simulation by implementation of a modified Gillespie's stochastic simulation algorithm (35) in MATLAB (MathWorks, Natick, MA). Cells are initialized at steady state protein numbers and then simulated for 12 "Monte-Carlo hours" to generate an equilibrium distribution at the initial TMG concentration. Results of the deterministic model were calculated by integrating Eqs. **1–3** using Euler's method with $\Delta t = 1$ min.

Because not all parameters used in the analysis are measured explicitly, it is important to test the model's behavior for robustness against parameter error (see supporting information). Although the theoretical predictions change quantitatively as individual parameters are varied, the important features are preserved through a wide range of parameter values.

We thank Drs. Mukund Thattai and Attila Becskei for fruitful discussions and Dr. Han Lim and Suzanne Komili for technical assistance and helpful discussions. This work was supported by National Institutes of Health Grant GM068957 and National Science Foundation Faculty Early Career Development Grant PHY-0094181. J.T.M. and D.M. are supported by National Science Foundation graduate research fellowships.

1. Rao, C. V., Wolf, D. M. & Arkin, A. P. (2002) *Nature* **420,** 231–237.
2. Kaern, M., Elston, T. C., Blake, W. J. & Collins, J. J. (2005) *Nat. Rev. Genet.* **6,** 451–464.
3. Raser, J. M. & O'Shea, E. K. (2005) *Science* **309,** 2010–2013.
4. Kussell, E., Kishony, R., Balaban, N. Q. & Leibler, S. (2005) *Genetics* **169,** 1807–1814.
5. Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L. & Leibler, S. (2004) *Science* **305,** 1622–1625.
6. Acar, M., Becskei, A. & van Oudenaarden, A. (2005) *Nature* **435,** 228–232.
7. Hasty, J., Pradines, J., Dolnik, M. & Collins, J. J. (2000) *Proc. Natl. Acad. Sci. USA* **97,** 2075–2080.
8. Pedraza, J. M. & van Oudenaarden, A. (2005) *Science* **307,** 1965–1969.
9. Arkin, A., Ross, J. & McAdams, H. H. (1998) *Genetics* **149,** 1633–1648.
10. Kepler, T. B. & Elston, T. C. (2001) *Biophys. J.* **81,** 3116–3136.
11. Thattai, M. & van Oudenaarden, A. (2001) *Proc. Natl. Acad. Sci. USA* **98,** 8614–8619.
12. Ozbudak, E. M., Thattai, M., Kurtser, I., Grossman, A. D. & van Oudenaarden, A. (2002) *Nat. Genet.* **31,** 69–73.
13. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. (2002) *Science* **297,** 1183–1186.
14. Swain, P. S., Elowitz, M. B. & Siggia, E. D. (2002) *Proc. Natl. Acad. Sci. USA* **99,** 12795–12800.
15. Blake, W. J., Kærn, M., Cantor, C. R. & Collins, J. J. (2003) *Nature* **422,** 633–637.
16. Paulsson, J. (2004) *Nature* **427,** 415–418.
17. Raser, J. M. & O'Shea, E. K. (2004) *Science* **304,** 1811–1814.
18. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. (2005) *Science* **307,** 1962–1965.
19. Hooshangi, S., Thiberge, S. & Weiss, R. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 3581–3586.
20. Becskei, A., Kaufmann, B. B. & van Oudenaarden, A. (2005) *Nat. Genet.* **37,** 937–944.
21. Colman-Lerner, A., Gordon, A., Serra, E., Chin, T., Resnekov, O., Endy, D., Pesce, C. G. & Brent, R. (2005) *Nature* **437,** 699–706.
22. Isaacs, F. J., Hasty, J., Cantor, C. R. & Collins, J. J. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 7714–7719.
23. Vilar, J. M., Guet, C. C. & Leibler, S. (2003) *J. Cell Biol.* **161,** 471–476.
24. Simpson, M. L., Cox, C. D. & Sayler, G. S. (2003) *Proc. Natl. Acad. Sci. USA* **100,** 4551–4556.
25. Novick, A. & Weiner, M. (1957) *Proc. Natl. Acad. Sci. USA* **43,** 553–566.
26. Maloney, P. C. & Rotman, B. (1973) *J. Mol. Biol.* **73,** 77–91.
27. Nobelmann, B. & Lengeler, J. W. (1996) *J. Bacteriol.* **178,** 6790–6795.
28. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & van Oudenaarden, A. (2004) *Nature* **427,** 737–740.
29. Chung, J. D. & Stephanopoulos, G. (1996) *Chem. Eng. Sci.* **51,** 1509–1521.
30. Wong, P., Gladney, S. & Keasling, J. D. (1997) *Biotechnol. Prog.* **13,** 132–143.
31. Yildirim, N. & Mackey, M. C. (2003) *Biophys. J.* **84,** 2841–2851.
32. Liu, J. Y., Miller, P. F., Willard, J. & Olson, E. R. (1999) *J. Biol. Chem.* **274,** 22977–22984.
33. Le, T. T., Harlepp, S., Guet, C. C., Dittmar, K., Emonet, T., Pan, T. & Cluzel, P. (2005) *Proc. Natl. Acad. Sci. USA* **102,** 9160–9164.
34. Muller-Hill, B., Crapo, L. & Gilbert, W. (1968) *Proc. Natl. Acad. Sci. USA* **59,** 1259–1264.
35. Gillespie, D. T. (1977) *J. Phys. Chem.* **81,** 2340–2361.

Table 3. Parameter robustness

Parameter	X_{high}	X_{low}
α	141.96	120.15
β	82.95	81.22
ρ	85.78	94.37
λ	177.18	120.88
N_{LacY}	73.62	91.61
N_{GFP}	82.42	94.70
N_{LacI}	75.12	89.13
b_{LacY}	83.29	63.02
b_{GFP}	90.80	71.45
b_{LacI}	70.27	74.07
$\tau_{1/2}$	95.61	119.71

Results of parameter robustness simulations where fractional error (X) is reported for changes in each individual parameter. For comparison, the cost calculated from two replicates of the same experiment is $X_{\text{Experimental}} = 45.50$ and the cost for parameters given by values in Table 1 is $X_0 = 67.02$.

Supporting Materials and Methods

Growth conditions and Media. Cells were grown at 37ºC in M9 minimal media with succinate as the main carbon source. Overnight cultures were prepared in the absence of TMG or in the presence of 100 μM TMG to yield populations of cells uninduced or fully induced for *lac* expression, respectively. Cells from the overnight cultures were diluted into fresh media containing intermediate TMG levels and maintained at low density ($0.001 < OD_{600} < 0.005$) to prevent TMG depletion throughout the experiment. At specified time points, a portion of each culture was removed and prepared for imaging; fresh, pre-warmed media with the appropriate TMG level was added to dilute the remaining culture so that the cell density of each imaged sample was the same ($OD_{600} = 0.005$) for all time points. Samples were concentrated and prepared for imaging by filtration, centrifugation, resuspension in 1.25 μl of the appropriate growth media, and placement on a microscope slide.

Bacterial Strains. Dynamic population distributions of *lac* promoter activity were gathered using *Escherichia coli* MUK21 (1), in which the *gfp* gene is placed under the control of a wild-type *lac* promoter and chromosomally inserted. Published steady-state distributions from the ERT113 strain (1) are further analyzed here for both noise measurements and partitioning analysis. Strain ERT113 was constructed by transforming MUK21 cells with a plasmid containing the red-fluorescent protein HcRed under the control of the *gat* promoter.

Fluorescence Microscopy. Measurements of GFP fluorescence in dynamic (pre-steady-state) cell populations were obtained using a Nikon TE300 microscope equipped with automatic stage and focus, and a Micromax 1024B CCD camera (Roper Scientific), all controlled by MetaMorph software (Universal Imaging). Steady-state measurements were gathered as previously described (1). Fluorescence values for cells are corrected by subtracting background fluorescence measured in a region of the field of view devoid of cells. Cell boundaries were determined by auto-thresholding phase contrast images, and GFP intensity was averaged over this area. Mean fluorescence levels are assumed to be representative of reporter concentration and are calculated by dividing the total intensity of the cell by the area in pixels of the cell in the phase contrast image. These numbers are then normalized so the induced population average of the mean fluorescence is 100 for both GFP and RFP.

Derivation of Deterministic Model. We modify a deterministic model for the lactose uptake network (1) as shown below:

$$\frac{R}{R_{\mathrm{T}}} = \frac{1}{1 + \frac{X^2}{X_0{}^2}} \quad \textbf{[6]}$$

$$\frac{dY}{dt} = \frac{N_{\mathrm{LacY}} / \tau_{\mathrm{y}}}{1 + \frac{R}{R_0}} - \frac{Y}{\tau_{\mathrm{y}}} \quad \textbf{[7]}$$

$$\frac{dG}{dt} = \frac{N_{\mathrm{GFP}} / \tau_{\mathrm{g}}}{1 + \frac{R}{R_0}} - \frac{G}{\tau_{\mathrm{g}}} \quad \textbf{[8]}$$

$$\frac{dX}{dt} = BY + \Lambda TMG_{\mathrm{external}} - \frac{X}{\tau_{\mathrm{x}}} \quad \textbf{[9]}$$

The fraction of active LacI tetramers (R) as a function of internal TMG (X) is modeled by a Hill function with Hill coefficient of two. (Eq. **6**). R_T is the total concentration of LacI tetramers, and X_0 represents the half-saturation point. The rate of production of LacY molecules (Y) and GFP molecules (G) in the presence of LacI (R) is also a Hill function with Hill coefficient of one (Eqs. **7** and **8**). Here R_0 is the half-saturation concentration of active LacI (R) while N_{LacY} and N_{GFP} are the equilibrium number of LacY and GFP molecules, respectively, in fully induced cells. B represents the active uptake of TMG per LacY molecule, while Λ represents the passive, LacY-independent uptake of external TMG (Eq. **9**). Y, G, and X are all assumed to undergo first-order decay with time constants τ_y, τ_g and τ_x, respectively. To derive the simplified equations (Eqs. **1-3**) shown in the main text, Eq. **6** is substituted into Eqs. **7** and **8**, and the following new parameters (Eq. **10**) are defined:

$$g \equiv \frac{\alpha}{N_{\mathrm{GFP}}} G$$

$$y \equiv \frac{\alpha}{N_{\mathrm{LacY}}} Y$$

$$x \equiv \frac{X}{X_0}$$

$$\rho \equiv \frac{R_{\mathrm{T}}}{R_0} + 1$$

$$\beta \equiv \frac{N_{\mathrm{LacY}}\tau_{\mathrm{x}}}{\alpha X_0} B$$

$$\lambda \equiv \frac{\tau_{\mathrm{x}}}{X_0} \Lambda \qquad \textbf{[10]}$$

Effect of Induction on Doubling Time. It has been suggested that in order to model stochastic transitions in the lactose uptake network, one must include effects of TMG induction on doubling time (2). We find that induction with TMG has no statistically significant change on growth rate; therefore, we ignore this effect in our model.

Estimating deterministic parameters: τ_x and λ. To estimate τ_x, the time constant of internal TMG decay, we examine a population of cells that has been grown for 24 hours in media containing 50 μM TMG, which is sufficient to force all cells into the fully induced state of high GFP expression. These cells were then transferred into media lacking TMG at $t = 0$. GFP measurements were made at $t = 4$, 6, 8 and 10 hours (Fig. 5*a*), and the average concentration of GFP was measured at each time point with an error estimated from the population standard deviation. Because there is no external TMG in the media, the cells will cease production of GFP when the internal concentration of TMG becomes sufficiently dilute. Given that GFP is not actively degraded, its level should fall exponentially based on the doubling time. Therefore, we fit an exponential curve to the mean GFP concentration at each time point, using the least-squares method. The resulting best-fit curve was extrapolated to $t = 0$ and used to determine the approximate time at which GFP production ceased (Fig. 5*c*). Our extrapolation indicates that GFP production in induced cells begins to decay within 10-20 minutes following removal of external TMG, suggesting that $\tau_x << \tau_{1/2}$. Thus, we assume Eq. **1** is in equilibrium when compared to Eqs. **2** and **3**.

We expect that uninduced cells switched into media with a high TMG concentration will transition quickly to the induced state, minimizing stochastic effects. In this situation the deterministic model should provide a reasonable estimate of this behavior at the population level. Therefore, we arrive at an estimate of λ, the rate of TMG leakage, by fitting the deterministic model to experimental data in which we observed cells transitioning from an uninduced state to a fully induced state. Cells initially grown in absence of TMG were switched into media containing 55 μM of TMG, and measurements were taken at $t = 1$, 2, 3 and 4 hours (Fig. 5*b*). The population average GFP level was determined from this distribution with error bars set by the standard deviation. The parameter λ was then varied from 0 to 1.0 in 0.0025 intervals, and the χ^2 between the predictions of the deterministic model and the population average was

calculated at each point (Fig. 5*d*). The best-fit value was found to be $\lambda = 0.06$, with an 80% confidence range of (0.03, 0.12). Because α, β, and ρ were determined with $\lambda = 0$, a non-zero value of λ will change the position of the monostable-bistable transition and in principle require the refitting of all other parameters. However, we find that the lower monostable-bistable boundary shifts only from 3.5 to 3.4 μM TMG, which is smaller than the precision at which the boundary was originally determined. The position of the upper monostable-bistable boundary is changed as well, but the behavior of the stochastic model still matches the experiments in this region indicating that refitting is unnecessary.

Measurement of Noise around Steady State. We show the main sources of noise in Fig. 1*b*. Each gene has a source of intrinsic noise, which is related to mRNA burst size, *b*, and protein number, *N*, by the relation $\eta_{\text{int}}^2 = \frac{b+1}{< N >}$. Because GFP expression and RFP expression share one source of noise, the correlation between the levels of these molecules will depend on this noise source and no others. To derive this relation, we can use the Langevin formalism in the same manner in which Pedraza *et al.* (3) have applied it to noise propagation in a synthetic gene cascade. Here we treat the rate of change of each molecule as having two components: continuous terms due to rates in the deterministic model and stochastic terms due to intrinsic noise. This is made more precise by writing the corresponding Langevin equations for the two extrinsic sources and the two reporters in induced cells, where LacI and LacY noise is not transmitted.

$$\begin{aligned} \tau\dot{E} &= 1 - E + \mu_{\text{E}} \\ \tau\dot{G} &= E\alpha_{\text{G}} - G + \mu_{\text{G}} \\ \tau\dot{R} &= E\alpha_{\text{R}} - R + \mu_{\text{R}} + \mu_{\text{P}} \end{aligned} \qquad \textbf{[11]}$$

Here E represents the collective effect of global noise factors (for example, CRP, ribosome and RNA polymerase concentrations) and is scaled to an equilibrium value of one. G and R are the number of GFP and RFP molecules, while α_G and α_R represent the equilibrium amount of GFP and RFP in induced cells. We assume that observed drops in the levels of these molecules result from cell division, so τ is the characteristic decay time due to dilution. The following relations define the noise terms accounting for intrinsic fluctuations of proteins.

$$\begin{aligned} &< \mu_{\text{G,R}} >= 0 \\ &< \mu_{\text{G,R}}(t)\mu_{\text{G,R}}(t+\Delta t) >= 2\tau\alpha_{\text{G,R}}(b_{\text{GFP,RFP}}+1)\delta(\Delta t) \end{aligned} \qquad \textbf{[12]}$$

We include a term μ_P to account for noise introduced by fluctuations in plasmid number, which we assume is uncorrelated with other sources of noise. Fluctuations of the global factors may be defined in a manner similar to that for intrinsic protein noise, but we do not *a priori* know the magnitude of these fluctuations:

$$< \mu_{\mathrm{E}} >= 0$$
$$< \mu_{\mathrm{E}}(t)\mu_{\mathrm{E}}(t+\Delta t) >=< \mu_{\mathrm{E}}^2 > \delta(\Delta t) \qquad \textbf{[13]}$$

We want to use Eq. **11** to derive the noise properties of our fluorescent reporters, which can be accomplished by calculating deviations from steady state values: $\delta E \equiv E - 1$, $\delta G \equiv G - \alpha_{\mathrm{G}}$, $\delta R \equiv R - \alpha_{\mathrm{R}}$. Substituting these calculations in Eq. **11**, Fourier transforming, squaring, and inverse Fourier transforming yields the following result for the second moment of δG and the correlation between δG and δR:

$$< \delta G^2 >= \alpha_{\mathrm{G}}(b_{\mathrm{GFP}} + 1) + \frac{\alpha_{\mathrm{G}}}{\alpha_{\mathrm{R}}} < \delta G \delta R > \qquad \textbf{[14]}$$

Eq. **14** relates the total fluctuations around steady state in GFP to a contribution from intrinsic fluctuations and a contribution from global fluctuations, meaning that we can separate the total noise in GFP ($\eta_{\mathrm{g-total}}$) into intrinsic ($\eta_{\mathrm{g-int}}$) and global (η_{global}) components by the relation

$$\eta_{\mathrm{g-total}}^2 = \eta_{\mathrm{g-int}}^2 + \eta_{\mathrm{global}}^2 \qquad \textbf{[15]}$$

Here, the individual noise contributions are related to the reporter fluctuations:

$$\eta_{\mathrm{g-total}}^2 = \frac{< \delta G^2 >}{\alpha_{\mathrm{G}}^2} = \frac{< G^2 > - < G >^2}{< G >^2}$$
$$\eta_{\mathrm{global}}^2 = \frac{< \delta G \delta R >}{\alpha_{\mathrm{G}}\alpha_{\mathrm{R}}} = \frac{< GR > - < G >< R >}{< G >< R >} \qquad \textbf{[16]}$$

The biochemical parameters α_{G} and b_{GFP} can be obtained from

$$\eta^2_{\text{g-int}} = \frac{b_{\text{GFP}} + 1}{\alpha_{\text{G}}} = \eta^2_{\text{g-total}} - \eta^2_{\text{global}} \qquad \textbf{[17]}$$

By assuming the mean fluorescence in a single cell is proportional to the concentration of molecules in that cell, we can measure total and extrinsic noise for a population using Eq. **16**, and the intrinsic noise is given by Eq. **17**. Surprisingly, we find that even in fully induced cells the fluctuations in GFP are coming almost entirely from intrinsic noise (Fig. 4*b*).

Measurement of the Number of Molecules. Decomposing the noise in a gene into intrinsic and extrinsic components still does not characterize fluctuations with sufficient detail to build a microscopic simulation; the numbers of relevant proteins in each cell is still needed. We estimate this number using a method similar to that introduced by Rosenfeld *et al.* (4), where GFP fluorescence is compared between dividing cells.

The process of cell division can be viewed as a binomial process where each molecule of GFP is randomly and independently assigned to one of the two daughter cells. Letting N_1 and N_2 be the number of molecules in the first and second daughter, respectively, we assume that the following relations fully determine the statistics of this process: $N_{\text{Pair}} = N_1 + N_2$, $< N_1 >= \frac{N_{\text{Pair}}}{2}$, and $< N_1^2 >= \frac{N_{\text{Pair}}(1 + N_{\text{Pair}})}{4}$, where averages are taken over the cell population. Furthermore, we assume that the mean fluorescence values, gfp_1 and gfp_2, in both daughter cells are directly proportional to the number of GFP molecules. Based on these assumptions, N_{GFP} can be estimated as shown below from the fluctuations of intensity between dividing cells without requiring details of photon flux or quantum efficiency.

$$\frac{< (gfp_1 - gfp_2)^2 >}{< gpf_1 >< gfp_2 >} = \frac{< (N_1 - N_2)^2 >}{< N_1 >< N_2 >} = \frac{4 < N_1^2 > -4N_{\text{Pair}} < N_1 > + N_{\text{Pair}}^2}{N_{\text{Pair}}^2 / 4} = \frac{4}{N_{\text{Pair}}} \qquad \textbf{[18]}$$

Eq. **18** relates the distribution of cellular fluorescence values in the daughter cells to the total number of molecules present in the mother. Because the fluorescence of an average cell is somewhere between the undivided cell (N_{Pair}) and the daughter cells ($N_{\text{Pair}}/2$), we assume that $N_{\text{GFP}} = 3/4N_{\text{Pair}}$. Finally we measure gfp_1 and gfp_2 for a population of 70 pairs of fully induced cells in 30 μM TMG and find that N_{GFP}=790±210 molecules with error bars estimated by bootstrapping.

Stochastic Model Details. Because mRNA production is a large source of noise, we build a model where noisy events are dominated by mRNA processes and use protein levels to determine the instantaneous state of the system. The three main events in the model are mRNA production/degradation, protein degradation, and global noise.

We assume the number of proteins produced from an individual mRNA to be chosen from an exponential distribution by treating the decay of mRNA as a random Poisson process and assuming that the number of proteins translated is proportional to the lifetime. Because the lifetime of an mRNA is very short in relation to the timescales associated with fluctuations in protein level ($\tau_{1/2}$) we condense the three events (production, translation and degradation) into a single 'burst' event. We model this process by production and immediate decay of an mRNA whose net effect was the addition of a random number of new proteins to the system. We quantify the rate at which these bursts occur by dividing the rate of protein production by the mean number of proteins produced from an mRNA.

We model the loss of protein levels by random Poisson decay of individual proteins at a rate commensurate with that caused by dilution. This adds noise to the system that is not inherent in cell growth and ignores noise due to the partitioning process. However, the noise difference between the decay and dilution processes should not be important because the noise from mRNA bursts is dominant in our experiments. This could become relevant for dynamics that are heavily influenced by decay processes when there are few if any mRNA burst events.

Global noise is included as a multiplicative term on the production of GFP and LacY. To include this effect, the random walk process *E(t)*, described in Eqs. **11** and **13**, is simulated and included in the relevant mRNA production rates. Here the value $<\mu_E^2>$ is chosen so that the simulated value of η_{global} matches that measured experimentally. The statistics on the process are thus given by $\langle E \rangle = 1$ and $\langle \delta E(t)\delta E(t+\Delta t)\rangle = 2\eta_{\text{global}}^2 \exp(-\Delta t/\tau_{1/2})$. This is then numerically simulated by a discrete random walk, and the mRNA production rates of relevant proteins are multiplied by the resultant factor $E(t)$.

Further Experimental Verification of Model. To verify that the model is functioning properly, growth of cells from both ON and OFF histories is simulated for the equivalent of 20 h in a range of external concentrations of TMG. As can be seen in Fig. 6 the cells largely remain in their original states through the shaded region, whereas transitions between the two states occur more frequently past the edges of this region. The hysteretic effect is similar to published measurements of single cells (1), indicating

that the model is capturing many of the essential properties of the biological network. The greater than expected transition rate from OFF-to-ON in the deterministically bistable region could be due to increased sensitivity to parameter error in this region, or an overestimation of LacY noise. In Fig. 7 we compare the model results to experimental data for cells grown in 6 μM and 9 μM TMG, both in the deterministically bistable region, and we find reasonable agreement between these data.

To further test our model in the bistable region, we calculate the OFF-to-ON transition rate when cells are switching stochastically. This can be done by fitting the equation $f_0 e^{-\gamma t}$ to the fraction of cells remaining in the OFF state at time t, $f_{OFF}(t)$. The results for the switching rate γ(TMG) from this fitting are shown in Fig. 8 for both experimental data and simulations in the range of 3 - 60 μM TMG. The transition rate for both curves is near or equal to zero until approximately 9 μM TMG, which coincides with the upper boundary of the deterministically bistable region. Above 9μM TMG, the model and experimental transition rates rise together as extracellular TMG increases.

Parameter Robustness. In order to quantify the sensitivity of the model's output to variations in the deterministic and noise parameters, we construct a cost function that compares two population distributions. The cost function should return small values for nearly identical distributions and large values for very different distributions. We set $H(G, T, t)$ to be the fraction of initially uninduced cells in the bin centered at G units of GFP fluorescence after t hours of growth in T μM TMG. This is similar to the curves shown in Fig. 2. We define a cost function, X, similar to a χ^2 error, and evaluate it on the logarithmic-normalized histograms:

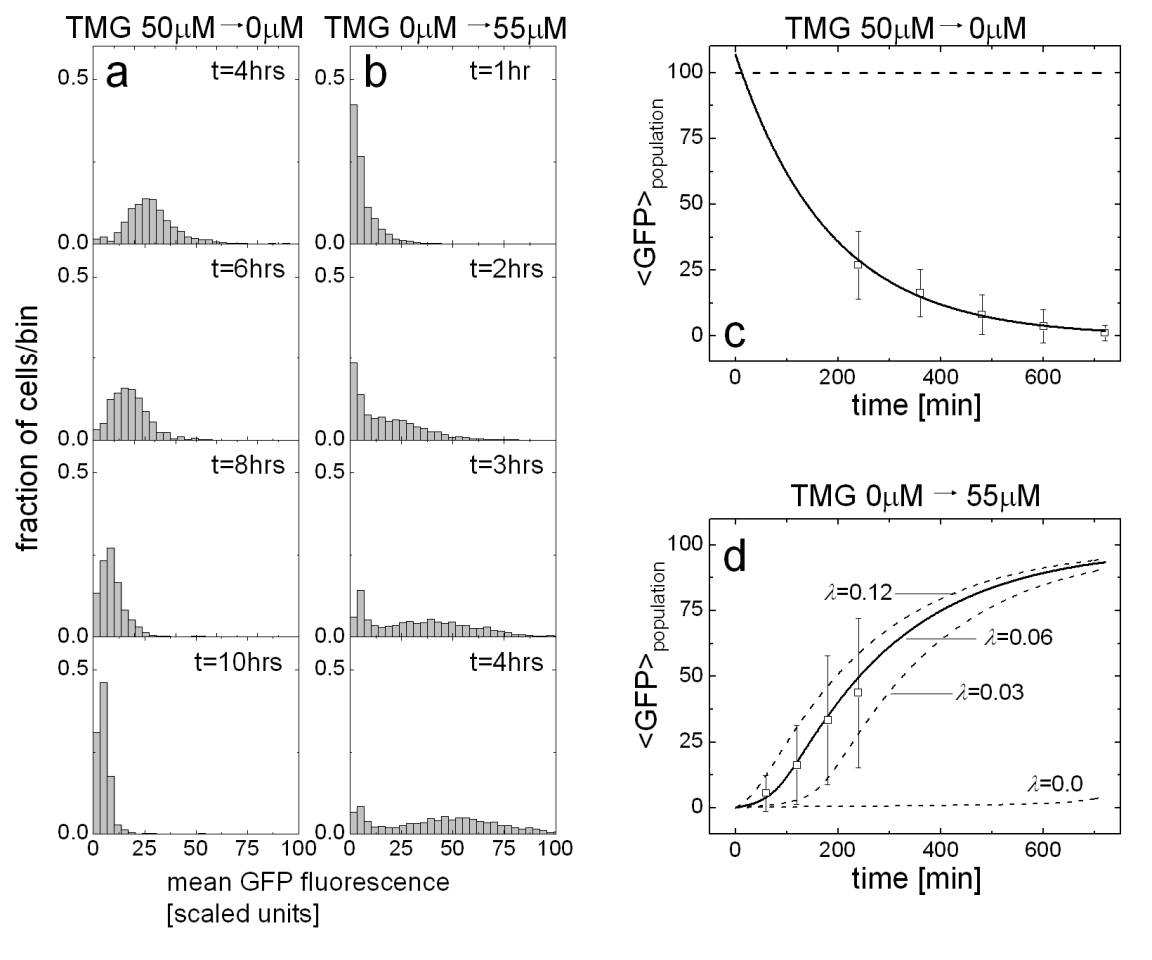
$$X = \Sigma_T \Sigma_t \Sigma_G \, log^2(H_{\text{experiment}}(G,T,t)/ H_{\text{model}}(G,T,t))$$

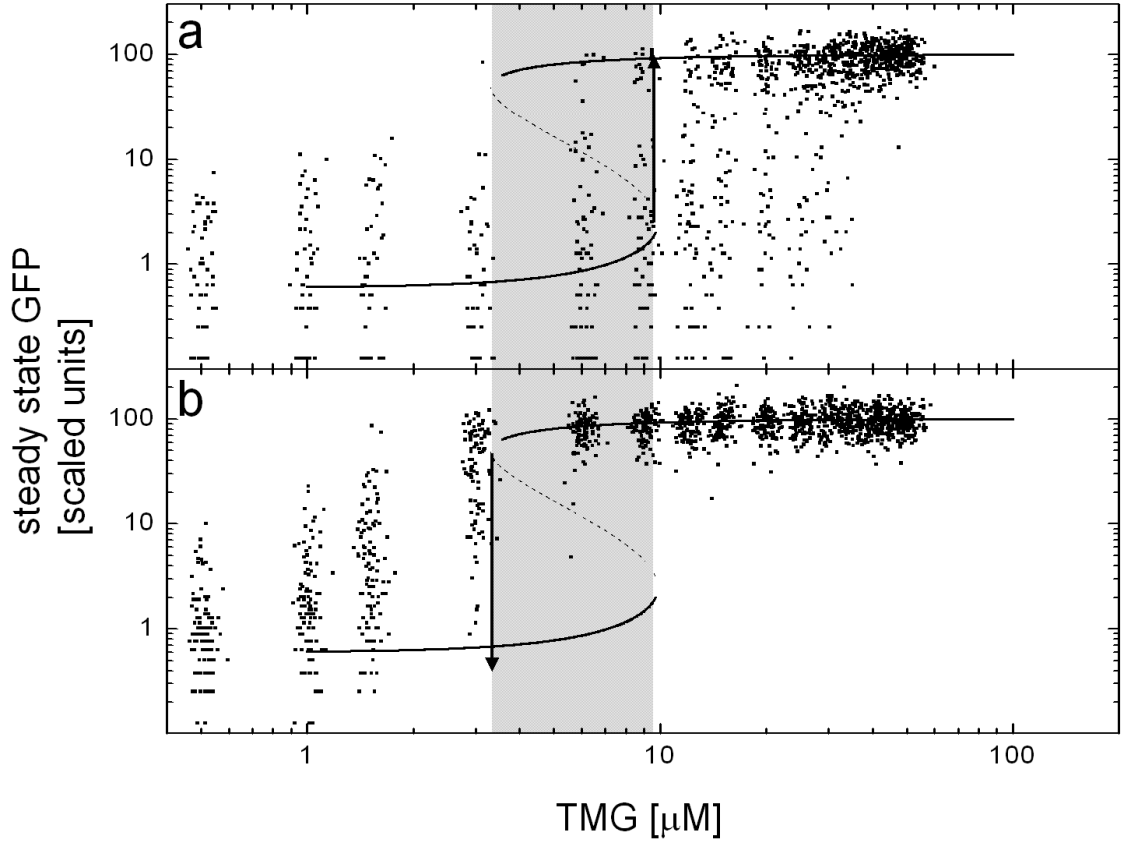
The sums are confined to the histograms shown in Figs. 3 *b-d*, and any terms with $H = 0$ were ignored in the sum. To estimate the sensitivity of the model to parameter error, each parameter was individually varied, and the cost between the model predictions and experimental measurements was calculated. We vary α, β, ρ, λ, N_{GFP} and b_{GFP} by the calculated errors shown in Table 2, while N_{LacY}, b_{LacY}, N_{LacI} and b_{LacI} are each varied by a factor of two because these parameter values were not measured directly. X_{high} and X_{low} represent the cost functions generated from simulations in which individual parameters are set to the upper and lower error boundaries, respectively. These calculations are shown for each parameter in Table 3. When all parameters are given by the values

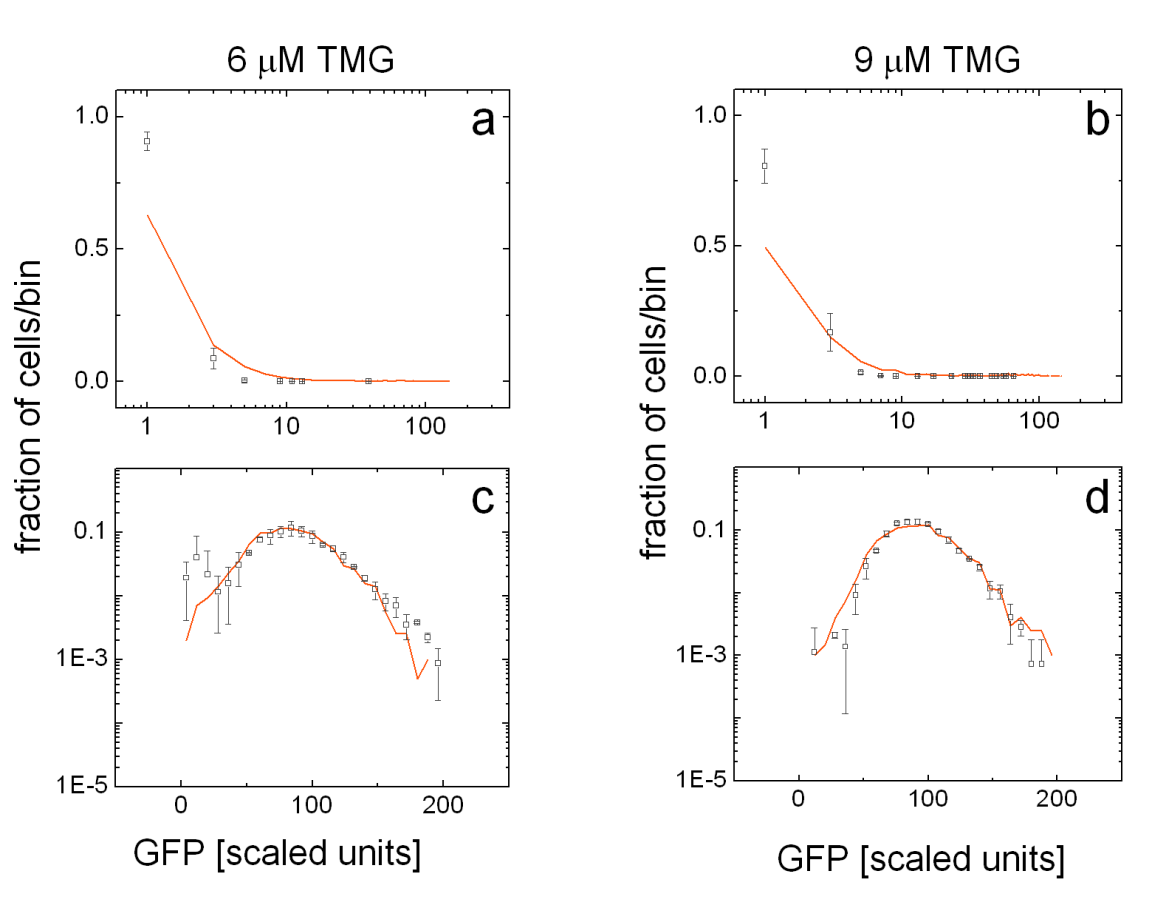
indicated in Table 1 the cost is found to be $X_0 = 67.02$, which is close to the cost between two histograms generated from replicates of the same experiment: $X_{\text{Experimental}} = 45.50$.

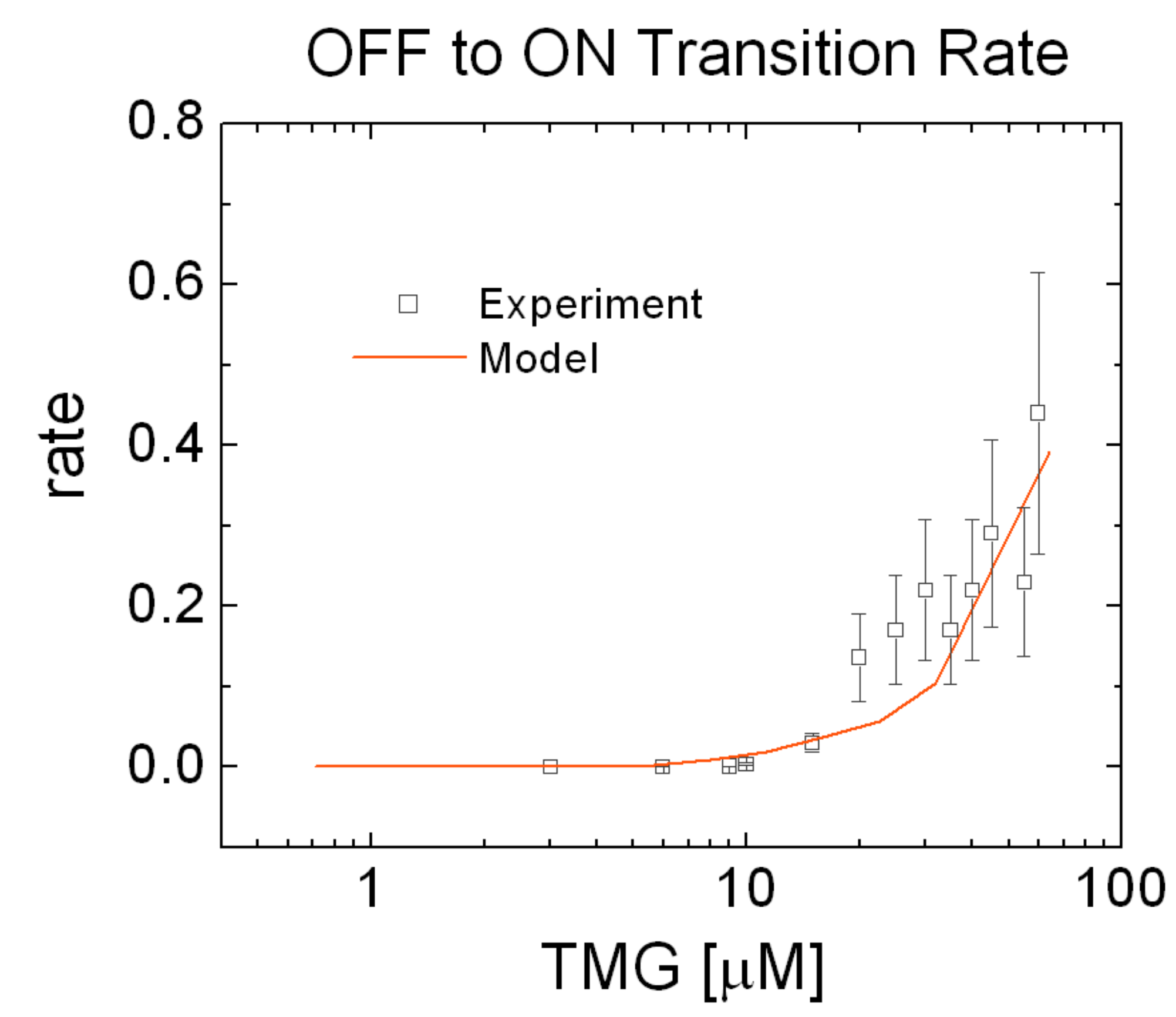
Error on the parameters α, λ, and $\tau_{1/2}$ has the greatest impact on the model's predictions when compared to experiments. Even though the highest value of X is more than twice as large as the lowest, the qualitative predictions remain similar throughout the range of parameters. Variations in each parameter slightly change the shape of the simulated histograms. For example, increased (decreased) values of λ created a higher (lower) rate of OFF to ON transitions. While the shapes of each peak in the bimodal histogram of transitioning cells remains similar, the relative magnitudes of the peaks are changed. For changes in α, however, the peak of OFF cells decays with a rate similar to experiments, but the simulated transitioning cells either produce GFP too quickly or too slowly for the histogram to closely match experiments. The noise parameters N and b seem to affect prediction accuracy the least. Decreasing b_{LacY} even causes better agreement than the value estimated through GFP noise, suggesting that noise in the LacY levels might have been over-estimated.

1. Ozbudak, E. M., Thattai, M., Lim, H. N., Shraiman, B. I. & Van Oudenaarden, A. (2004) *Nature* **427,** 737-740.
2. Vilar, J. M., Guet, C. C. & Leibler, S. (2003) *J Cell Biol* **161,** 471-476.
3. Pedraza, J. M. & van Oudenaarden, A. (2005) *Science* **307,** 1965-1969.
4. Rosenfeld, N., Young, J. W., Alon, U., Swain, P. S. & Elowitz, M. B. (2005) *Science* **307,** 1962-1965.


TMG 50μM→0μM
TMG 0μM →55μM
a
b
t=4hrs
t=1hr
t=6hrs
t=2hrs
t=8hrs
t=3hrs
t=10hrs
t=4hrs
fraction of cells/bin
mean GFP fluorescence
[scaled units]
TMG 50μM→ 0μM
<GFP>population
time [min]
c
TMG 0μM → 55μM
d
λ=0.12
λ=0.06
λ=0.03
λ=0.0


a
b
100
10
1
1
10
100
steady state GFP
[scaled units]
TMG [μM]


6 μM TMG
9 μM TMG
a
b
c
d
fraction of cells/bin
fraction of cells/bin
1.0
0.5
0.0
1
10
100
0.1
1E-3
1E-5
0
100
200
GFP [scaled units]
GFP [scaled units]


OFF to ON Transition Rate
0.8
0.6
0.4
0.2
0.0
rate
Experiment
Model
1
10
100
TMG [μM]