Noisy information processing through transcriptional regulation

Eric Libby*, Theodore J. Perkins†, and Peter S. Swain*‡

*Centre for Nonlinear Dynamics, Department of Physiology, McGill University, 3655 Promenade Sir William Osler, Montreal, QC, Canada H3G 1Y6; and †McGill Centre for Bioinformatics, McGill University, 3775 University Street, Montreal, QC, Canada H3A 2B4

Edited by Robert H. Austin, Princeton University, Princeton, NJ, and approved February 23, 2007 (received for review October 10, 2006)

Cells must respond to environmental changes to remain viable, yet the information they receive is often noisy. Through a biochemical implementation of Bayes's rule, we show that genetic networks can act as inference modules, inferring from intracellular conditions the likely state of the extracellular environment and regulating gene expression appropriately. By considering a two-state environment, either poor or rich in nutrients, we show that promoter occupancy is proportional to the (posterior) probability of the high nutrient state given current intracellular information. We demonstrate that single-gene networks inferring and responding to a high environmental state infer best when negatively controlled, and those inferring and responding to a low environmental state infer best when positively controlled. Our interpretation is supported by experimental data from the *lac* operon and should provide a basis for both understanding more complex cellular decision-making and designing synthetic inference circuits.

biochemical networks | systems biology | Bayesian inference

For cells to interact with their environment, the DNA and regulatory machinery, which are intracellular, require information from the cell surface. This information is conveyed through gene and protein networks and is transferred via biochemical reactions that are potentially significantly stochastic (1–4). Stochastic fluctuations will undermine both signal detection and transduction. Cells are therefore confronted with the task of predicting the state of the extracellular environment from noisy and potentially unreliable intracellular signals. For example, a bacterium must decide from intracellular levels of a nutrient whether or not the nutrient is sufficiently abundant extracellularly to express the appropriate catabolic enzymes. Similarly, a smooth muscle cell must decide from concentrations of second messengers whether or not extracellular hormone levels are high enough to warrant contracting.

Here, we consider if, and how, it is possible for biochemical networks to correctly infer properties of the extracellular environment based on noisy, intracellular signals. Suppose that the cell should respond under high concentrations of an extracellular molecule. Suppose further that the concentration of an intracellular signaling molecule is related to the concentration of the extracellular molecule through a signal transduction mechanism. A simple inference network could establish a concentration threshold for the intracellular molecule. Only if the molecule is above threshold is the extracellular concentration judged to be high enough for a cellular response. This network performs poorly, however, in fluctuating extracellular and intracellular environments. First, fluctuations lead to input molecules crossing threshold even when the state of the environment is unchanged. Second, a threshold scheme cannot specify the degree of certainty in the inference, which may be important for the ultimate response. For example, a bacterium may express a catabolic operon once the degree of certainty in high extracellular levels of a particular nutrient reaches 40%, but it may only shut down other catabolic operons once the degree of certainty is larger, say 80%.

The method of Bayesian inference both accounts for fluctuations and gives a degree of uncertainty in predictions (5). We postulate that the cellular regulatory machinery may have evolved to perform Bayesian inference on some intracellular inputs. Typically, a cellular decision has two levels: first, predicting the state of the environment; second, choosing the appropriate response. At this second level, the expected costs must be compared with expected benefits (6). Although Bayesian theory can handle both problems, we focus here on the first: classification of the local environment.

As an example, consider a bacterium with a nutrient scavenging operon that encodes enzymes to import and catabolize a sugar (Fig. 1 *A* and *B*). Suppose the environment can be in one of two states: a high or a low sugar state, for example, the high- and low-lactose environments of the small intestine (7). The intracellular concentration of the sugar depends on the extracellular state, although in a stochastic fashion. To optimize growth, the bacterium must predict the extracellular state from intracellular sugar because expressing the operon involves a significant metabolic cost (6, 8). Let S be the intracellular sugar level at a particular time. We denote the probability (i.e., the fraction of time) that there are S intracellular sugar molecules given that the environment is in the low sugar state as $P(S|\text{low})$. Similarly, we denote the probability that there is S intracellular sugar molecules given that the environment is in the high sugar state by $P(S|\text{high})$. If fluctuations are negligible, these two distributions will be sharply peaked functions of S, and they will be broader as fluctuations become significant.

The bacterium must determine the probability that its extracellular environment is in a high sugar state based on levels of intracellular sugar. This probability is denoted $P(\text{high}|S)$. A Bayesian approach assumes that some information about the long-term probable states of the environment is known. This information could be simply that the environment is expected to be in one of two states, either a low or a high sugar state, and that each state is *a priori* equally likely. In one particular environment (for example, the soil), though, a low sugar state may occur more often on the long term. The *a priori* probability for this state will then be higher. Such *a priori*, or prior, probabilities are denoted $P(\text{high})$ and $P(\text{low})$. Once sugar enters the cell, the *a priori* probabilities are updated based on the levels of sugar detected. The more intracellular sugar, the larger the predicted probability of the environment being in the high sugar state (and the smaller the corresponding probability of the low sugar state). This *a posteriori* probability of the high state is $P(\text{high}|S)$. It is referred

Author contributions: T.J.P. and P.S.S. designed research; E.L. performed research; E.L., T.J.P., and P.S.S. analyzed data; and E.L., T.J.P., and P.S.S. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Abbreviation: IPTG, isopropyl β-D-thiogalactoside.

‡To whom correspondence should be addressed. E-mail: swain@cnd.mcgill.ca.

This article contains supporting information online at www.pnas.org/cgi/content/full/0608963104/DC1.

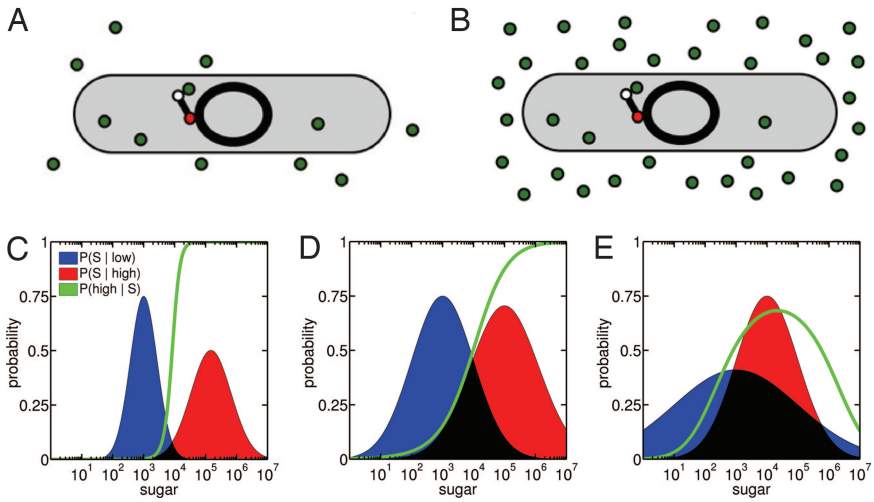

Fig. 1. A two-state classifier problem and its Bayesian solution, the posterior probability. A cell must infer from intracellular concentrations of a nutrient or signaling molecule (green circles) whether the molecule is in high or low concentrations in the extracellular environment. (*A* and *B*) Fluctuations in the environment and molecule detection and transport can lead to similar intracellular concentrations of the molecule for different extracellular conditions. The cellular decision-making machinery, shown as a simple genetic network, must decide from intracellular information the probable state of the extracellular environment. (*C*) Two distributions for intracellular numbers of a sugar molecule: the low sugar state is in blue, the high sugar state is in red. For an intracellular sugar level *S*, the green curve is the posterior (predicted) probability that the extracellular state is the high sugar state, *P*(high|*S*). (*D*) For two intracellular distributions that overlap substantially, the posterior probability for the high sugar state transitions gradually from low to high values. (*E*) The posterior probability, *P*(high|*S*), need not be monotonic. The low sugar state is more probable at both low and high intracellular sugar, and *P*(high|*S*) goes through a maximum.

to as the posterior (predicted) probability of the high state given intracellular sugar *S*.

Bayes's rule states explicitly how the prior probabilities are correctly updated to their posterior values for the levels of sugar detected (9) (see *Materials and Methods*):

$$P(\text{high}|S) = \frac{P(S|\text{high})\,P(\text{high})}{P(S|\text{low})\,P(\text{low}) + P(S|\text{high})\,P(\text{high})}. \quad [1]$$

Intuitively, the more likely a particular intracellular *S* is in the high extracellular state compared with the low extracellular state [the greater *P*(*S*|high) is compared with *P*(*S*|low)], the higher the posterior probability of a high state environment. For simplicity, we will assume that the environment is *a priori* equally likely to be in either state: *P*(high) = *P*(low) = 1/2. The prior probabilities then play no mathematical role in Eq. **1**. Often the posterior distribution, *P*(high|*S*), is a sigmoidal curve. Fig. 1*C* shows two distributions for numbers of sugar molecules: a distribution for a low extracellular sugar state (in blue) and a distribution for a high extracellular sugar state (in red). The corresponding posterior probability curve is shown in green in Fig. 1*C*. If the intracellular sugar level, *S*, is low, there is a high predicted probability that the extracellular state is low, with the converse holding for high intracellular sugar levels. In an intermediate range of *S*, lying in the overlap between the two state distributions, *P*(high|*S*) switches from low probability to high probability. When fluctuations are more significant and the overlap between the two distributions is greater, the transition is more gradual (Fig. 1*D*). The posterior probability need not always be sigmoidal: Fig. 1*E* shows a long-tailed distribution for the low sugar state that results in a nonmonotonic posterior curve.

We will argue that a single gene can make probabilistic inferences about extracellular states through a biochemical implementation of Bayes's rule. By tuning the kinetic rates of the system, the promoter efficacy, the fraction of time the promoter is capable of initiating transcription, can match the posterior probability of high extracellular sugar. Consider a negatively controlled operon. We view the repressors controlling the gene as detectors that monitor intracellular sugar levels. Repressors thermally flip back and forth between two allosteric forms (10): one DNA binding and the other non-DNA binding. As each repressor diffuses in the cytosol, it samples intracellular sugar. At low sugar levels, the DNA binding form of the repressor is stable, and the operon is not expressed. At high sugar levels, the non-DNA binding form is stable, leading to expression. Repressor binding sites on the promoter "read" the allosteric form of cytosolic repressors and control transcription. Promoter efficacy is therefore a readout of the number of non-DNA binding repressors, which, in turn, are a readout of sugar levels.

Cis-Regulatory Regions as Inference Modules

We tested the ability of different regulatory mechanisms to classify a two-state environment. We considered 18 different networks (Fig. 2 *A–C*): regulation can be positive or negative, transcription factor can allosterically bind either one, two, or four sugar molecules, and promoters can be one of three different types. Network input is the number of sugar molecules, which range from zero to ≈2,000 times the number of transcription factors. Network output is promoter efficacy (i.e., promoter bound by an activator for positive control and free of repressor for negative control). Rather than specialize to particular sugar distributions for the high and the low states, we generated 50 different pairs of lognormal distributions for *S*. Each pair corresponded to a different inference problem and had a different, but always sigmoidal, posterior probability. We fit the kinetic rates of each network to minimize the squared error between promoter efficacy and *P*(high|*S*) as a function of *S* for each of the 50 posteriors (see *Materials and Methods*). A network that fits this collection of posterior curves well has a network architecture able to solve a variety of (two state) inference problems; it is an inference module.

Networks with higher cooperativity, either through the ability of transcription factor to allosterically bind sugar or cooperative binding of transcription factors to DNA, perform best (Fig. 2 *D* and *E*). A genetic inference system with low cooperativity is unable to generate a promoter efficacy curve that switches

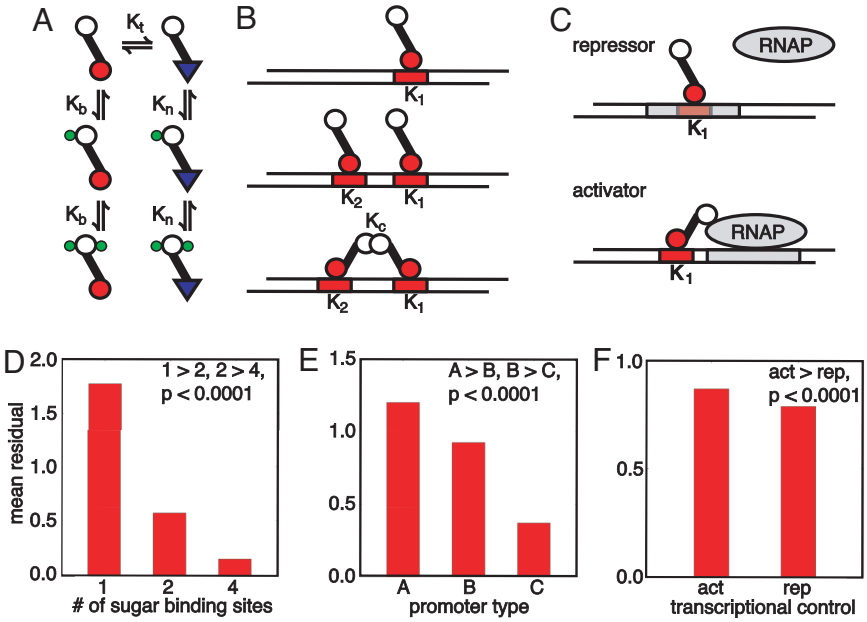

Fig. 2. A comparison of different regulatory mechanisms for solving the two-state discrimination problem; highly cooperative, negatively controlled genetic networks perform the most accurate inference. (*A*) The Monod–Changeux–Wyman model of an allosteric transcription factor. Association constants are denoted by *K*s. The protein flips between DNA binding (red circles) and non-DNA binding forms (blue triangles). If $K_b \gg K_n$, sugar stabilizes the DNA binding state. Conversely, if $K_n \gg K_b$, the non-DNA binding state is stabilized. Two sugar binding sites are shown, but we also test models with one and four binding sites. (*B*) We consider three different promoters: type A, one active operator site (*Top*); type B, two active operator sites, but with no cooperative binding between transcription factors (*Middle*); and type C, one active and one inactive operator with cooperative transcription factor binding (*Bottom*). (*C*) Transcription can by regulated either negatively, via repressors that obstruct RNA polymerase (RNAP) binding, or positively, via activators that help stabilize RNAP binding. The RNAP binding site (sigma site) is shown in gray, operators in red. (*D*) Mean residuals (a high residual implies a poor fit) from fits to 50 different posterior probabilities for the models grouped by the numbers of sugars bound by transcription factor. Models with four transcription binding sites perform the best inference (*P* value for one model type consistently performing better than the other is given; see *SI Appendix*). (*E*) Mean residuals for models grouped by promoter type. Cooperative promoters perform best (type C). (*F*) Mean residuals for models grouped by their mode of transcriptional control. Repressors perform better than activators (for >70% of the fits, corresponding to a *P* value substantially $<10^{-4}$).

sharply with S (10). These models thus perform poorly on those inference problems with distinct sugar distributions and therefore strongly sigmoidal posterior probabilities (compare the posterior probabilities for Fig. 1 *C* and *D*).

Less intuitively, negatively controlled inference systems perform significantly better than positively controlled systems (Fig. 2*F*). Positively controlled systems are less able to exploit cooperativity. Activators should bind DNA as sugar levels rise. Consequently, $K_b \gg K_n$ in Fig. 2*A*. For low sugar, the posterior probability is close to zero (Fig. 1 *C* and *D*), and no activators at all should bind DNA. Therefore K_b must be small, and the more activators present, the smaller K_b must be. As $K_b \gg K_n$, both K_b and K_n are small: there is weak sugar binding, and cooperative binding occurs only at high sugar levels. Contrarily, in a negatively controlled system, $K_n \gg K_b$, so that sugar lifts repressor off DNA. For low sugar, just one repressor must bind DNA to maintain a low promoter efficacy. More repressors allow K_b to be smaller, giving greater, not less, flexibility in K_n. Altering K_t, the equilibrium between the DNA and non-DNA binding forms in the absence of sugar and can partly offset the inherent frustration in the activator system, but not completely (Fig. 2*F*). Therefore, negatively controlled promoters are best able to tune promoter efficacy to track $P(\text{high}|S)$.

Although negatively controlled systems can better match their promoter efficacy to $P(\text{high}|S)$ than positively controlled systems, the opposite holds for matching $P(\text{low}|S)$. This posterior probability satisfies $P(\text{low}|S) = 1 - P(\text{high}|S)$ and so has the opposite behavior to $P(\text{high}|S)$. The argument given above is reversed. Thus, for systems that respond to a low state of the environment, positive control gives the best inference.

Fig. 2 demonstrates that model genetic networks can perform inference, with equilibrium promoter efficacy tracking posterior probability; Fig. 3 shows that inference can occur in real time in noisy environments. For the two sugar distributions in Fig. 1*C*, we chose the activator and repressor networks that best fit the posterior probability of the high sugar state. We performed a stochastic simulation of each of these networks by using the best-fit parameters, and let the environment change from a low to a high and back to a low sugar state. In each state, we sampled from the appropriate sugar distribution, mimicking intracellular fluctuations, and producing a time series of intracellular sugar (Fig. 3*A*). For each sugar level, there is a different posterior probability of the high extracellular sugar state (Fig. 1*C*). This instantaneous posterior probability is shown in Fig. 3*B*. Most often, $P(\text{high}|S)$ is very low (near zero) or very high (near one). It should be compared with the response of each network, measured by their promoter efficacies (Fig. 3 *C* and *D*). The promoter efficacy of the repressor network (Fig. 3*C*) and the activator network (Fig. 3*D*) closely follow the instantaneous posterior probability, although the activator network underestimates the probability of the high sugar state. A quantitative measure of the goodness of fit of each promoter efficacy to $P(\text{high}|S)$ shows that repressor performs more than twice as well as activator [see supporting information (SI) *Appendix*].

Inference in the *lac* Operon

Viewing networks as inference modules gives additional interpretations of *in vivo* behavior. For example, Setty *et al.* (11) measured the transcription rate of the *lac* operon in *Escherichia coli* as a function of two inputs: isopropyl β-D-thiogalactoside (IPTG), an analogue of lactose, and cAMP. Traditionally, transcription of the

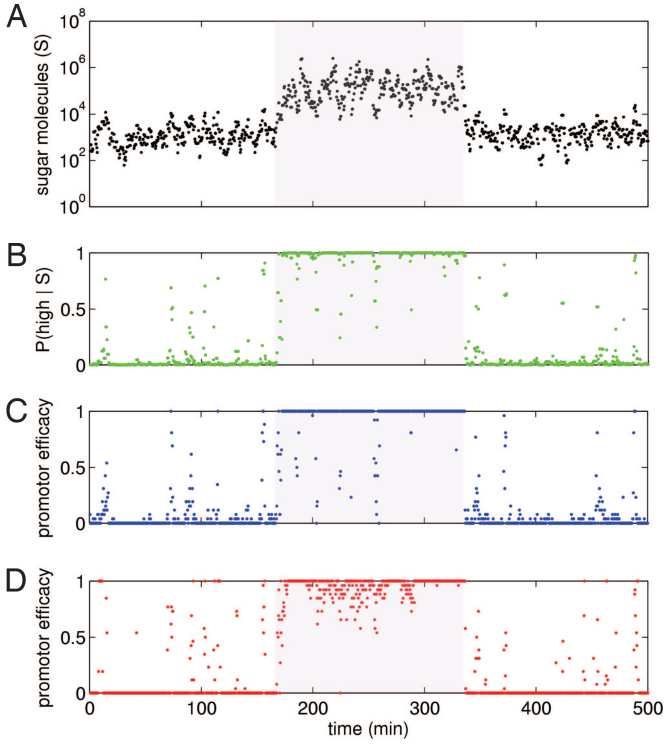

Fig. 3. Two-state inference by simulated genetic networks. (*A*) A time series of intracellular sugar molecules as the extracellular environment moves from a low to a high (shaded region) and back to a low sugar state. Histograms of the intracellular sugar distributions are shown in Fig. 1*C*. Sugar was sampled every 25 s. In the low state, mean sugar numbers are $\approx 10^3$; in the high state, mean sugar numbers are $\approx 10^5$. (*B*) The instantaneous posterior probability of the high sugar state, $P(\text{high}|S)$, for the particular sugar level existing at the current time point. Posterior probability points come from the green curve in Fig. 1*C*. (*C*) The average promoter efficacy for the best repressor network of Fig. 2, four sugar binding sites on the repressor and promoter type C. The actual promoter efficacy is either zero (promoter bound by repressor) or one (promoter not bound). An average over the 25-s period chosen to sample the sugar is shown. (*D*) The average promoter efficacy for the best activator network of Fig. 2, again four sugar binding sites on the activator and promoter type C. Simulation details are in *SI Appendix*.

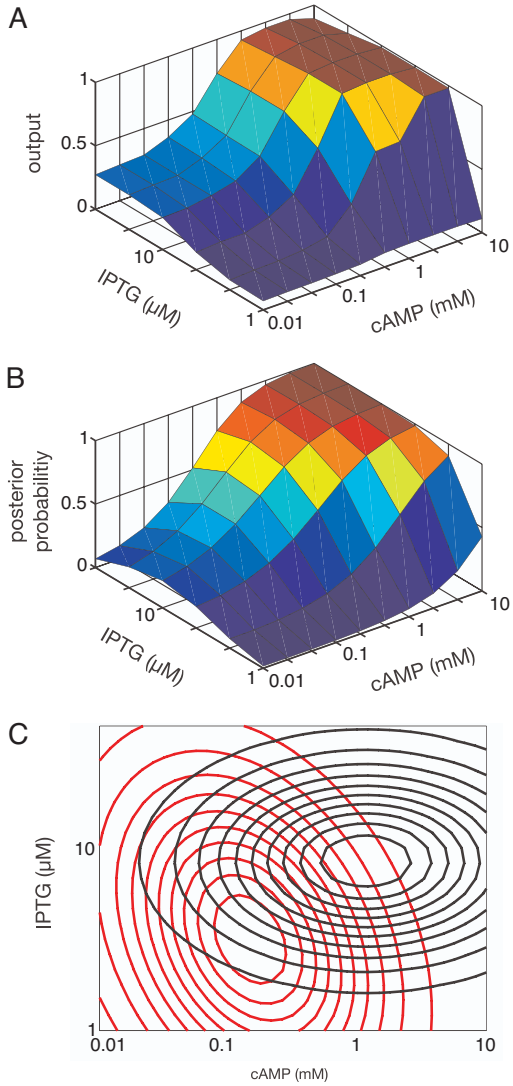

Fig. 4. Inference by the *lac* operon in *E. coli*. (*A*) Observed transcriptional output (transcription rate) as a function of extracellular concentrations of IPTG and cAMP (both log-scaled), normalized to range from zero to one (data from ref. 11). (*B*) Posterior probability, fit to the data in *A*, that the environment is in a high state given the concentrations of IPTG and cAMP. (*C*) A possible two-state model for *E. coli*'s view of its extracellular environment. The low state is in red (peak at approximately 3 μM IPTG and 0.2 mM cAMP), and the high state is in black (peak at 8 μM IPTG and 1.2 mM cAMP). Both states are described by bivariate lognormal distributions.

lac operon is described as being "on" in the presence of sufficient cAMP and sufficient lactose, i.e., its cis-regulatory region performs a logical "AND" on the two inputs (12). Setty *et al.* found more complex behavior: with enough IPTG, there is significant transcription at low cAMP, and transcription increases smoothly, rather than in a switch-like fashion, as cAMP increases (Fig. 4*A*). The shape of this surface can be explained if the *lac* operon has evolved to solve a two-state inference problem. The high state corresponds to a state where the *lac* operon should be expressed, an extracellular environment rich in lactose and poor in glucose, resulting in both high intracellular lactose and cAMP [cAMP concentrations are inversely proportional to glucose levels (13)]. The low state, where the *lac* operon should not be repressed, corresponds to an extracellular environment poor in lactose and rich in glucose. We interpret S in Eq. **1** as the set of two variables: intracellular IPTG and cAMP concentrations (see *Materials and Methods*). Assuming bivariate lognormal distributions for IPTG and cAMP in each state, we fit the parameters of the distributions so that the posterior probability, $P(\text{high}|S)$, matches the data of Fig. 4*A* (Fig. 4*B*). Two lognormal distributions that generate this posterior are shown in Fig. 4*C*. [Note that the axes represent measured extracellular levels, which are assumed to be proportional to intracellular levels (11).] The *lac* transcription rate is explained well by a two-state model in which mean intracellular levels of IPTG are approximately three times higher in the high state than in the low state and cAMP levels are 10 times higher.

Discussion

We have argued that a single gene through allosteric control and its cis-regulatory region can statistically infer the state of the extracellular environment from intracellular inputs. Cis-regulatory regions are often considered to perform logical operations on their input, allowing gene expression only under a particular combination of inputs (14, 15). Such a view has been especially successful in understanding development (16), where gene expression occurs in an ordered manner. Cell behavior need not, however, follow a predetermined pattern, and in these cases a cell that infers the state of its environment may have an evolutionary advantage. A genetic network, or more generally a biochemical network, that performs inference allows the cell to

optimally interpret fluctuating inputs. Expression of the *lac* operon is a possible example, but inference is also likely to occur in signal transduction networks. Although we have emphasized the sigmoidal character of the posterior probability, networks that perform Bayesian inference need not have a sigmoidal output. Fig. 1*E* shows two sugar distributions that produce a biphasic posterior probability. Such behavior has been reported, for example, in the *E. coli gal* operon (17), and is hard to justify within a logic gate description.

We predict that a positively controlled genetic inference module is more likely to infer the probability of the environment being in a low state and that a negatively controlled system is more likely to infer the probability of the environment being in a high state. For example, the cAMP receptor protein in *E. coli* is an activator and promotes high promoter efficacy of the *lac* operon when glucose levels are low; LacI is a repressor and promotes high promoter efficacy when lactose levels are high (12). This bias is expected to be stronger for networks with less cooperativity.

Although we have focused on a single estimate of the probability of the extracellular state, cells might be expected to perform long-term integration of noisy signals. Such integration could occur by changing the prior probabilities of the high and low states. For example, an *E. coli* previously exposed to lactose has a higher concentration of lactose permease in its cell membrane than one not exposed (18). This greater permease concentration may reflect an increase in the prior probability of the high extracellular lactose state, i.e., $P(\text{high}) > P(\text{low})$. Eq. **1** then predicts a sigmoidal response that favors the high state: the posterior probability curve is shifted toward lower sugar levels. This change mimics the change expected in promoter efficacy of the *lac* operon: higher permease concentrations lead to gene expression (higher promoter efficacy) at lower extracellular lactose levels because lactose more efficiently enters the cell.

In our framework, the output of different networks are distinct functions of their input because each network is solving a different inference problem. For example, if the intracellular distributions of the two extracellular states strongly overlap, a repressor may have a high allosteric constant (K_t in Fig. 2*A*) to give a more sigmoidal promoter efficacy curve, reflecting the steep posterior probability. The promoter efficacy curve is most sensitive, however, to the inducer binding affinity (K_n for repressors and K_b for activators). Its sensitivity is more than three times higher than the next most sensitive parameter (K_t) (see *SI Appendix*). If the extracellular environment substantially changes, leading to a new inference problem, the most efficient way to evolve to the new posterior probability is to modify the sugar binding affinity. This modification has the benefit of preserving the connectivities of preexisting genetic networks.

Cellular inference need not follow the simple two-state classifier model proposed here. Multistate classifiers and real-time averaging methods are more appropriate for some problems. Nevertheless, given the prevalence of sigmoidally responding biochemical networks (19), the two-state classifier, whose solution is often a sigmoidal posterior probability, may be an essential component of many inference and decision-making networks in cells. Interpreting biochemical networks as inference modules may be an important step for both unraveling cellular behavior and designing selective, synthetic gene circuits.

Materials and Methods

Modeling Genetic Networks. We use the Monod–Wyman–Changeux model (10) to describe allosteric transcription factors. We assume that both the total amount of sugar and the total amount of transcription factors are conserved. Given these values, we numerically solve for the amount of free sugar and the total amount of transcription factor in the DNA binding state, irrespective of the number of sugars each individual transcription factor has bound (see *SI Appendix*).

To calculate promoter efficacies, we follow a statistical mechanics approach (20) to describe the equilibrium occupancies of the different states of the promoters of Fig. 2*B* (see *SI Appendix*).

Comparison of the Models as Bayesian Classifiers. To test the ability of the models to implement a Bayesian classifier, we fit each model to the posterior probabilities for 50 different two-state classification problems. For each problem, we generated two sugar distributions corresponding to a low and a high sugar state. From these distributions, we calculated the posterior probability of being in the high state for each concentration of sugar S:

$$P(\text{high}|S) = \frac{P(S|\text{high})\,P(\text{high})}{P(S)}. \qquad [2]$$

We can rewrite the expression for the probability of a sugar concentration as:

$$\begin{aligned} P(S) &= \sum_{\text{states}} P(S|\text{state})P(\text{state}) \\ &= P(S|\text{high})P(\text{high}) + P(S|\text{low})P(\text{low}) \end{aligned} \qquad [3]$$

to derive Eq. **1**. For simplicity, we assume equal priors; allowing unequal prior probabilities for the two states does not change our results.

We considered two-state classification problems generated by Poisson, normal, and lognormal distributions of sugar. The results of Fig. 2 *D–F* are for lognormal distributions, but are qualitatively the same independent of the distribution type chosen. The probability $P(S|\text{state})$ in Eq. **1** is therefore:

$$P(S|\text{state}_i) = \frac{e^{\frac{-(\ln S - \mu_i)^2}{2\sigma_i^2}}}{\sqrt{2\pi}\sigma_i S}, \qquad [4]$$

where $i = 1$ for the low state and $i = 2$ for the high state. Each state has a different μ_i and σ_i, which define the mean and standard deviation in log space of the distribution. We chose 50 posterior probability curves that best gave a range of different inference problems (see *SI Appendix*).

We used a least-square fit to score how well a model matches the posterior probability of the high state. To fit we use an interior-reflective Newton method (lsqnonlin in Matlab, Mathworks, Natwick, MA). Each posterior probability curve generated has 100 points (evenly spaced in log space), and we fit all 18 models to each curve 500 times with different initial conditions, for a total of 450,000 fits. The P values for the residual comparisons were computed by using a Wilcoxon two-sided signed rank test (signrank in Matlab). For each fit, we calculated the difference in the residual for a particular pair of models. The null hypothesis was that these differences came from a distribution with median zero.

Stochastic Simulation. We simulated both a repressor and an activator model. We chose a posterior probability from the 50 used in the fitting (the posterior of Fig. 1*C*) and the repressor and activator model that fit it best (parameters are given in *SI Appendix*). The selected repressor and activator models both have four sugar binding sites and promoter type C in Fig. 2*B*. To generate a relatively smooth time series of sugar levels, we used a Markov chain Monte Carlo method (5) to produce fluctuating, dependent samples of sugar from the appropriate distribution in Fig. 1*C*. For each sugar sample, the cytosolic sugar levels were changed to the new sampled value. A stochastic simulation of the genetic network was then run for a fixed time interval of 25 s by using the Gillespie

algorithm (21) (results for different time intervals are given in *SI Appendix*). A new sugar sample was then taken and the simulation of the genetic network run again. The average value of the promoter efficacy during each simulation run is shown in Fig. 3 *C* and *D*.

Fitting a Posterior Probability to the Transcription Rate of the *lac* Operon. We fit the data of Fig. 4*A* to Eq. **1** where each state is characterized by two variables: s_1 corresponding to the logarithm of the IPTG concentration and s_2 corresponding to the logarithm of the cAMP concentration. $P(S|\text{high})$ is then a bivariate normal distribution:

$$P(S|\text{high}) \sim \frac{1}{\sqrt{\det(\sigma)}} \exp\left(-\frac{1}{2}\sum_{i,j=1}^{2}(s_i-\mu_i)\sigma_{ij}^{-1}(s_j-\mu_j)\right) \quad [5]$$

with μ_1 the mean of s_1, μ_2 the mean of s_2, and σ the covariance matrix of s_1 and s_2, all for the high state. A similar set of parameters is needed to describe the low state. The problem of fitting Eq. **1** to a given posterior probability surface is degenerate: different sets of parameters can result in the same posterior surface (see *SI Appendix*). However, we can identify a unique posterior probability surface that best fits the *lac* operon data (Fig. 4*B*) along with the family of two-state discrimination problems that generate the posterior surface. Fig. 4*C* shows one example of this family.

We thank Uri Alon, Julie Desbarats, Michael Elowitz, Leon Glass, Terry Hebert, Moises Santillan, and particularly Sharad Ramanathan for helpful comments and Yaki Setty and Uri Alon for supplying the data for Fig. 4*A*. P.S.S. holds a Tier II Canada Research Chair. P.S.S. and T.J.P. are supported by the Natural Sciences and Engineering Research Council and the Mathematics of Information Technology and Complex Systems National Centre of Excellence.

1. Ozbudak EM, Thattai M, Kurtser I, Grossman AD, van Oudenaarden A (2002) *Nat Genet* 31:69–73.
2. Elowitz MB, Levine AJ, Siggia ED, Swain PS (2002) *Science* 297:1183–1186.
3. Blake WJ, Kaern M, Cantor CR, Collins JJ (2003) *Nature* 422:633–637.
4. Raser JM, O'Shea EK (2004) *Science* 304:1811–1814.
5. Mackay DJC (2003) *Information Theory, Inference, and Learning Algorithms* (Cambridge Univ Press, New York).
6. Dekel E, Alon U (2005) *Nature* 436:588–592.
7. Savageau MA (1998) *Genetics* 149:1677–1691.
8. Koch AL (1983) *J Mol Evol* 19:455–462.
9. Cox RT (1946) *Am J Phys* 14:1–13.
10. Monod J, Wyman J, Changeux JP (1965) *J Mol Biol* 12:88–118.
11. Setty Y, Mayo AE, Surette MG, Alon U (2003) *Proc Natl Acad Sci USA* 100:7702–7707.
12. Ptashne M, Gann A (2002) *Genes and Signals* (Cold Spring Harbor Lab Press, Cold Spring Harbor, NY).
13. Makman RS, Sutherland EW (1965) *J Biol Chem* 240:1309–1314.
14. Thomas R (1973) *J Theor Biol* 42:563–585.
15. Glass L (1975) *J Chem Phys* 63:1325–1335.
16. Yuh CH, Bolouri H, Davidson EH (1998) *Science* 279:1896–1902.
17. Mukhopadhyay J, Sur R, Parrack P (1999) *FEBS Lett* 453:215–218.
18. Novick A, Weiner M (1957) *Proc Natl Acad Sci USA* 43:553–566.
19. Ferrell JE (2002) *Curr Opin Cell Biol* 14:140–148.
20. Shea MA, Ackers GK (1985) *J Mol Biol* 181:211–230.
21. Gillespie DT (1977) *J Phys Chem* 81:2340–2361.

SI Appendix

Modelling Genetic Networks

Allosteric Model of Transcription Factors

We model transcription factors as allosteric molecules having two states: a DNA-binding state (B) and a non-DNA binding state (N). Following Monod-Wyman-Changeux (1), the presence of sugar causes a shift in the time the transcription factor spends in each state (Fig. 5). Sugar can bind to either the B or the N form of the transcription factor, but does so with a different binding affinity (K_b for the DNA-binding state and K_n for the non-DNA-binding state). Only when unbound by sugar can the transcription factor change between its two states. The reaction describing this change has an equilibrium constant of K_t. If $K_b \gg K_n$, sugar preferentially binds to the B state. By binding to the transcription factor, sugar converts B_0 molecules into B_r molecules, more so than N_0 molecules into N_r molecules (where the subscript r denotes that r sugar molecules are bound). The reaction between B_0 and N_0 is no longer at equilibrium and more N molecules convert to B molecules while this equilibrium is restored. The population of transcription factors as a whole is now more in the stronger sugar-binding B state, and so again more B than N molecules are likely to bind sugar. This positive feedback means that the number of transcription factors in the B state can be a highly non-linear function of the number of sugar molecules (1). If $K_n \gg K_b$ the opposite behaviour occurs, and sugar drives the transcription factors into the non-DNA binding N state.

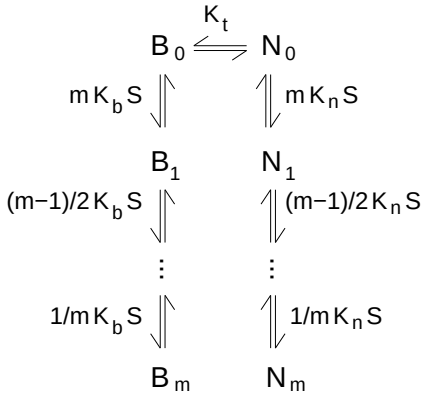

Fig. 5. An allosteric transcription factor that binds sugar S and exists in a DNA-binding state (B) and a non-DNA binding state (N). Each sugar binding site is assumed identical, and the subscripts denote the number of bound sugar molecules. Consequently, the basic equilibrium association constants for sugar binding, K_b and K_n, are altered by the ratio of the number of sites available for binding sugar (which increase the forward rate of the reaction) to the number of bound sugars (which increase the backward rate).

We assume that both the total amount of sugar and the total amount of transcription

factors are conserved:

$$S_{\text{tot}} = S + \sum_{r=0}^{m}(rN_r + rB_r) \quad [6]$$

$$T_{\text{tot}} = \sum_{r=0}^{m}(N_r + B_r), \quad [7]$$

where m is the number of sugar binding sites. Following ref. 1, we assume that each reaction in Fig. 5 is at equilibrium:

$$\begin{aligned} K_t N_0 &= B_0 \\ mK_b S B_0 &= B_1 \\ (m-1)K_b S B_1 &= 2B_2 \\ \vdots &\quad \vdots \\ K_n S N_{m-1} &= mN_m. \end{aligned} \quad [8]$$

Each equilibrium concentration can be solved in terms of N_0, the amount of transcription factor in the non-DNA binding state unbound by sugar:

$$\begin{aligned} N_r &= \binom{m}{r}(K_n S)^r N_0 \\ B_r &= \binom{m}{r}(K_b S)^r K_t N_0. \end{aligned} \quad [9]$$

Using these expressions and carrying out the summations in Eqs. **6** and **7** with the binomial theorem gives:

$$S_{\text{tot}} = S + N_0 m S\left[K_n(1+K_n S)^{m-1} + K_t K_b (1+K_b S)^{m-1}\right] \quad [10]$$

$$T_{\text{tot}} = N_0\left[(1+K_n S)^m + K_t(1+K_b S)^m\right]. \quad [11]$$

For a given S_{tot} and T_{tot} and the equilibrium association constants K_t, K_b, and K_n, we numerically solve Eqs. **10** and **11** for the amount of free sugar, S, and for N_0. We can therefore calculate the total amount of transcription factor in the non-DNA binding state, $N = N_0(1+K_n S)^m$, and the total amount in the DNA-binding state, $B = N_0 K_t (1+K_b S)^m$.

Promoter Models

We consider three different models of the promoter (Fig. 2*B* and *C*). The type A model has just one operator site. The type B model has two operators: a transcription factor at either operator prevents or initiates transcription independently. The final model, type C, has two operators but only one is sufficiently close to the RNA polymerase binding site to directly affect transcription. Nevertheless, a transcription factor bound to the inactive operator can stabilize a transcription factor bound to the active operator. We denote the fraction of time that the promoter is able to initiate transcription at equilibrium as promoter efficacy, P_{eff}.

We follow Shea and Ackers (2) to calculate the occupancy of the promoter at equilibrium. For example, for a negatively controlled type A promoter, which has just one binding site for a repressor, we consider the promoter existing in two states: P_1, bound by repressor, and P_0, not bound by repressor. If K_1 is the association constant for repressor binding and B is the number of repressors that are able to bind DNA, then $P_1 = K_1 B P_0$. The promoter is conserved: $P_0 + P_1 = 1$, if there is only one copy of the promoter. Combining these two equations implies that the promoter efficacy, P_0, obeys $P_0 = 1/(1 + K_1 B)$. We solve for the promoter efficacy for more complicated promoters similarly.

For a negatively controlled system, P_{eff} is the equilibrium fraction of promoter free from repressor. For the different promoter models:

type A,

$$P_{\text{eff}} = \frac{1}{1 + K_1 B} \qquad [12]$$

type B,

$$P_{\text{eff}} = \frac{1}{1 + (K_1 + K_2)B + K_1 K_2 B^2} \qquad [13]$$

type C,

$$P_{\text{eff}} = \frac{1 + K_2 B}{1 + (K_1 + K_2)B + \frac{1}{2}(K_1 K_2 + K_1 K_2 K_c)B^2}, \qquad [14]$$

where B is the total amount of transcription factor in the DNA-binding form, K_1 and K_2 are association constants for transcription factor binding to the two operator sites, and K_c determines the degree of cooperativity between two interacting, DNA bound transcription factors.

For positively controlled systems, P_{eff} is the equilibrium fraction of promoter bound by activator. For

type A,

$$P_{\text{eff}} = \frac{K_1 B}{1 + K_1 B} \qquad [15]$$

type B,

$$P_{\text{eff}} = \frac{(K_1 + K_2)B + K_1 K_2 B^2}{1 + (K_1 + K_2)B + K_1 K_2 B^2} \qquad [16]$$

type C,

$$P_{\text{eff}} = \frac{K_1 B + \frac{1}{2}(K_1 K_2 + K_1 K_2 K_c)B^2}{1 + (K_1 + K_2)B + \frac{1}{2}(K_1 K_2 + K_1 K_2 K_c)B^2}. \qquad [17]$$

Note when $K_c = 1$, that is, no cooperative interaction between the transcription factors, the type C models do not reduce to the type B models because only one operator is active for type C, whereas both are active for type B.

Comparison of the Models as Bayesian Classifiers

Generating the Posterior Probabilities

To generate a set of two-state classification problems, we assumed that each state can be described by a lognormal distribution:

$$P(S|\text{state}_i) = \frac{\mathrm{e}^{\frac{-(\ln S-\mu_i)^2}{2\sigma_i^2}}}{\sqrt{2\pi}\sigma_i S}. \qquad [18]$$

The low state has a sugar distribution with mean μ_1 and standard deviation σ_1; the high state has a mean μ_2 and standard deviation σ_2. We choose μ_1 to be either 1, 3, or 5; μ_2 to be either 5.1, 6.6, 8.1, or 9.6; σ_1 to be either 0.4, 0.5, 0.6, 0.7, 0.8, or 0.9; and σ_2 to be either 1, 1.25, 1.5, 1.75, or 2. All possible combinations of these parameters were considered, and we chose 50 pairs of distributions that best gave a range of different posterior probabilities (Fig. 6).

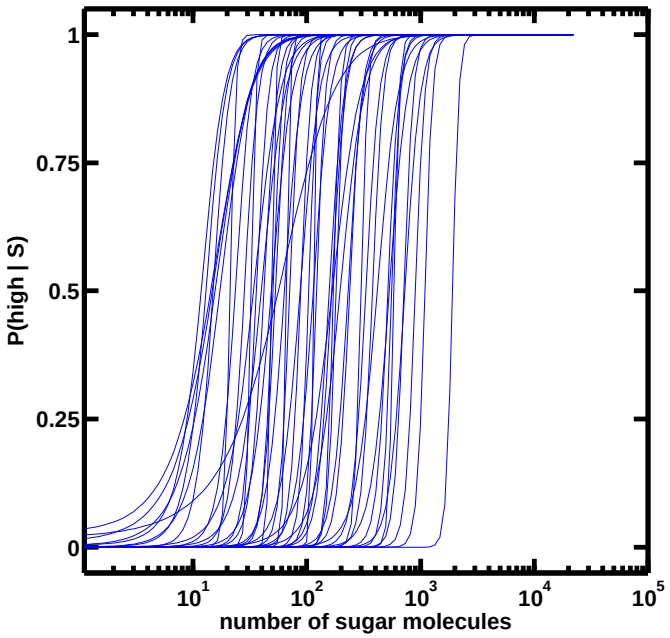

Fig. 6. The collection of posterior probabilities that were generated as solutions of lognormal two-state classification problems and used to compare the different genetic models of Fig. 2 as Bayesian classifiers.

Fitting the Models to the Posterior Probabilities

We used a least-square fit to score how well a model matches the posterior probability of the high state. The residuals plotted in Fig. 2 are the minimum value of the sum of squares:

$$\sum_i^n \Big[P(S_i|\text{high}) - P_{\text{eff}}(S_i,\lambda)\Big]^2, \qquad [19]$$

where we have n sugar levels S_i leading to n points on the posterior probability curve, $P(S|\text{high})$, we are trying to fit, and $P_{\text{eff}}(S,\lambda)$ is the model prediction for the promoter efficacy. This prediction is a function of the set of parameters λ: K_r, K_n, K_b, K_1, and K_2 and K_c depending on the promoter type. The minimum value of Eq. **19** occurs at the best-fit set of parameters λ. To ensure that the fitting algorithm considers only non-negative parameters, we define new variables for each parameter in log space. For example, $\kappa_1 = \log(K_1)$, and therefore can range over positive and negative values (3).

To correctly compare the ability of different models to fit a data set, models with more parameters should be penalized because they have more freedom to match the data. Typical methods are the Bayes Information Criterion (BIC) (4) and the Laplace method for model selection (5). With both of these techniques to compare the different models, the results of Fig. 2*D–F* were qualitatively unchanged. For the Laplace method, we need the maximum likelihood of the data (the 100 posterior probability points in our case) given the model. For each parameter, the maximum likelihood is penalized by a term that is determined by the error in the best fit value of the parameter and by its prior (5). We use:

$$\left(\sum_i^n \Big[P(S_i|\text{high}) - P_{\text{eff}}(S_i,\lambda)\Big]^2\right)^{-\frac{(n-1)}{2}} \qquad \textbf{[20]}$$

for the likelihood. This distribution results from assuming that the data have normally distributed errors with zero mean and any non-negative standard deviation (5). It is maximized when the sum of squares residual, Eq. **19**, is minimized.

Parameter Sensitivity

The sensitivities of the parameters were calculated as the mean log gain sensitivities (6) of the promoter efficacy. For parameter p_j, the sensitivity, χ_j, is

$$\chi_j = \left\langle \frac{\partial \log P_{\text{eff}}}{\partial \log p_j} \right\rangle, \qquad \textbf{[21]}$$

where the angled brackets denote an average over all sugar concentrations. We analytically calculated the $\partial \log P_{\text{eff}}/\partial \log p_j$ derivative as an implicit function of $\partial N_0/\partial p_j$ and $\partial S/\partial p_j$ by differentiating the promoter efficacy, such as Eq. **15** for example. We calculated these last two derivatives by differentiating Eqs. **10** and **11** with respect to p_j and numerically solving the resulting equations. Sensitivity values are given in Table 1.

Robustness of the Best-Fit Parameters

The fits of the promoter efficacy to the posterior probability curves are robust to changes in all but two of the parameters specifying each model. To investigate this robustness, we considered the model that best fit the posterior probability curves of Fig. 6. This model is transcriptionally controlled by a repressor that has four sugar binding sites. We varied each parameter individually and calculated the average change in the sum of squares residual, Eq.

Table 1. Parameter sensitivities for repressor and activator models

	Repressor model	**Activator model**
K_t	0.07	0.10
K_n	0.21	0.004
K_b	0.004	0.20
K_1	0.07	0.07
K_2	0.02	0.04
K_c	0.06	0.04

Parameters are defined in Fig. 2.

19, over all the posterior curves. The results shown in Fig. 7 reflect Table 1: the fit is only significantly sensitive to K_n, the sugar binding affinity for the non-DNA binding form of the repressor, and to a much lesser extent to K_t, the affinity describing transitions between the DNA- and non-DNA binding forms. Nevertheless, the sum of squares residual is so small for this model that the promoter efficacy curves behave like the posterior probability of Fig. $1C$ even if the residual is increased 5,000-fold (see Fig. 7 inset). We comment on possible implications of the high sensitivity to K_n in the text.

Stochastic Simulation

Table 2. Parameter values for the simulation shown in Fig. 3.

	Repressor model	**Activator model**
K_r	1.27 (10 s)	1.61×10^6 (10 s)
K_n	9.45×10^5 (10 s)	3.04×10^4 (10 s)
K_b	233 (10 s)	1.33×10^6 (10 s)
K_1	3.41×10^6 (0.1 s)	3.62×10^6 (0.1 s)
K_2	3.34×10^{10} (0.1 s)	1.51×10^9 (0.1 s)
K_c	88.3 (10 s)	219 (10 s)

These values are association affinities and are the best-fit values of the networks to the posterior probability of Fig. $1C$. Each association affinity is dimensionless because we simulate with numbers of molecules rather than concentrations. Shown in brackets is the corresponding dissociation rate. These rates, which are not given by a fit to $P(\text{high}|S)$, were chosen so that the network would respond in a reasonable time to changes in sugar levels.

To confirm that genetic networks can perform inference in real time with a noisy sugar source, we simulated both a repressor and an activator model with fluctuating sugar levels. We chose the posterior of Fig. $1C$ and the repressor and activator model that fit it best (parameters are given in Table 2).

To generate a relatively smooth time series of sugar levels, we used a Markov chain Monte Carlo method (5) to sample from the distributions in Fig. $1C$ (the Metropolis algorithm with a Gaussian trial distribution). We sample from the low distribution for 10^4 s, then from

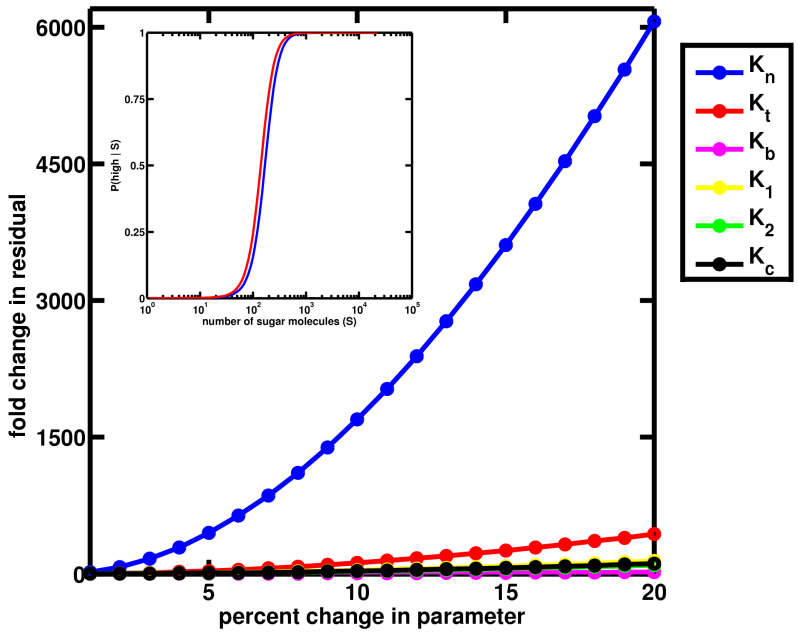

Fig. 7. Robustness of the sum of squares fit to systematic perturbations in individual model parameters away from their best-fit values. The model that best fits the posterior probabilities of Fig. 6 is shown: this model has promoter type C and is negatively regulated by a repressor with four sugar binding sites. Parameters are defined in Fig. 2. (Inset) An example of the promoter efficacy curves where K_n is changed by 20%. The curves are very similar despite the residual for the upper red curve being $\approx$ 5,000-fold larger than the residual of the original blue curve.

the high distribution for 10^4 s, and the again from the low distribution for another 10^4 s. For each sugar sample, the cytosolic sugar levels in the simulation are changed to the new sampled value. A stochastic simulation of the genetic network is then run for a fixed time interval (either 5, 10, 25, 50, or 100 seconds) by using the Gibson-Bruck version (7) of the Gillespie algorithm (8). The probability of a given reaction per unit time is equal to the product of the kinetic rate for the reaction and the number of potential reactants present. The time steps between reactions obey a Markov process. The cytosolic sugar level is then resampled by using the Markov chain Monte Carlo method and another Gillespie simulation run for this new level of sugar. The promoter efficacy plotted in Fig. 3 is the average promoter efficacy generated during each run of the Gillespie algorithm. Simulations start with one DNA molecule, 25 transcription factors in the DNA binding state and 25 transcription factors in the non DNA binding state.

For each choice of sugar sampling interval, we compared the performance of the two networks (Fig. 3*C* and *D*) to the instantaneous posterior probability (Fig. 3*B*). The comparison was scored by measuring the mean over time of the absolute difference between promoter efficacy and the instantaneous posterior probability. The results are shown in Table 3. Both networks perform better as the sugar sampling interval increases. As the time period grows over which the promoter efficacy is averaged, the average more closely matches the posterior

probability of Fig. 2*C* (for a long sampling period, the promoter efficacy will match the posterior probability almost perfectly because we use the best fit parameters for the simulation).

Negatively controlled networks consistently performed better because the network is better able to use its cooperativity (see the argument given in the text).

Table 3. Comparison scores of the mean absolute difference between the promoter efficacy and the instantaneous posterior probability of the high sugar state

Sampling interval, s	**Repressor model**	**Activator model**
5	7.0×10^{-2}	12.5×10^{-2}
10	3.3×10^{-2}	6.3×10^{-2}
25	3.4×10^{-2}	7.4×10^{-2}
50	2.8×10^{-2}	5.0×10^{-2}
100	2.5×10^{-2}	4.4×10^{-2}

Each score is the average from five simulation runs. A score of zero implies the the promoter efficacy exactly follows the instantaneous posterior probability. The sampling interval is the time between the samples of sugar used to generate the sugar time series.

Fitting a Posterior Surface to the Transcription Rate of the *lac* Operon

The Inverse Gaussian Classification Problem

A bivariate, or two dimensional, Gaussian distribution is a function of a vector (s_1, s_2) and is specified by a mean vector (μ_1, μ_2) and a 2×2 covariance matrix $\boldsymbol{\sigma}$. For example, μ_1 is the mean of the s_1 variable and σ_{11} its variance. $P(s_1, s_2)$ obeys:

$$P(s_1, s_2) \sim \frac{1}{\sqrt{\det(\boldsymbol{\sigma})}} \exp\left(-\frac{1}{2}\sum_{i,j}(s_i - \mu_i)\sigma_{ij}^{-1}(s_j - \mu_j)\right), \qquad \textbf{[22]}$$

where $\boldsymbol{\sigma}^{-1}$ is the matrix inverse of $\boldsymbol{\sigma}$.

A two-state, bivariate Gaussian classification problem is described by the prior probabilities of the two states, $P(\mathrm{II})$ and $P(\mathrm{I}) = 1 - P(\mathrm{II})$; the mean $\boldsymbol{\mu}^{\mathrm{I}}$ and covariance matrix $\boldsymbol{\sigma}^{\mathrm{I}}$ for s_1 and s_2 for state I; and the mean $\boldsymbol{\mu}^{\mathrm{II}}$ and covariance matrix $\boldsymbol{\sigma}^{\mathrm{II}}$ for s_1 and s_2 for state II. Given an observation of s_1 and of s_2, the posterior probability of state II is:

$$\begin{aligned} P(\mathrm{II}|s_1, s_2) &= \frac{P(s_1, s_2|\mathrm{II})P(\mathrm{II})}{P(s_1, s_2|\mathrm{I})P(\mathrm{I}) + P(s_1, s_2|\mathrm{II})P(\mathrm{II})} \\ &= \left(1 + \frac{P(s_1, s_2|\mathrm{I})P(\mathrm{I})}{P(s_1, s_2|\mathrm{II})P(\mathrm{II})}\right)^{-1}. \end{aligned} \qquad \textbf{[23]}$$

Inserting Eq. **22** in Eq. **23** gives:

$$P(\mathrm{II}|s_1, s_2) = \left(1 + \sqrt{\frac{\det(\boldsymbol{\sigma}^{\mathrm{II}})}{\det(\boldsymbol{\sigma}^{\mathrm{I}})}} \times \frac{\exp(-\frac{1}{2}\sum_{i,j}(s_i - \mu_i^{\mathrm{I}})(\sigma^{\mathrm{I}})_{ij}^{-1}(s_j - \mu_j^{\mathrm{I}}))}{\exp(-\frac{1}{2}\sum_{i,j}(s_i - \mu_i^{\mathrm{II}})(\sigma^{\mathrm{II}})_{ij}^{-1}(s_j - \mu_j^{\mathrm{II}}))} \times \frac{1 - P(\mathrm{II})}{P(\mathrm{II})}\right)^{-1}. \qquad \textbf{[24]}$$

From the posterior surface $P(\mathrm{II}|s_1, s_2)$, we would like to recover the parameters of the classification problem: $P(\mathrm{II})$, $\boldsymbol{\mu}^{\mathrm{I}}$, $\boldsymbol{\sigma}^{\mathrm{I}}$, $\boldsymbol{\mu}^{\mathrm{II}}$, and $\boldsymbol{\sigma}^{\mathrm{II}}$. This recovery is degenerate: different sets of parameters can result in the same posterior surface. With a little algebra, Eq. **24** can be reduced to the general form:

$$P(\mathrm{II}|s_1, s_2) = \left[1 + \exp(c_0 + c_1 s_1 + c_2 s_2 + c_3 s_1 s_2 + c_4 s_1^2 + c_5 s_2^2)\right]^{-1}, \qquad \textbf{[25]}$$

where the c_i depend on the parameters $P(\mathrm{II})$, $\boldsymbol{\mu}^{\mathrm{I}}$, $\boldsymbol{\sigma}^{\mathrm{I}}$, $\boldsymbol{\mu}^{\mathrm{II}}$, and $\boldsymbol{\sigma}^{\mathrm{II}}$. Although these parameters have 11 degrees of freedom [$P(\mathrm{II})$, two each for vectors $\boldsymbol{\mu}^{\mathrm{I}}$ and $\boldsymbol{\mu}^{\mathrm{II}}$, and three each for the covariance matrices $\boldsymbol{\sigma}^{\mathrm{I}}$ and $\boldsymbol{\sigma}^{\mathrm{II}}$], the posterior surface only has six degrees of freedom. The parameters therefore have five unrecoverable degrees of freedom.

Fitting the Transcription Rate Surface from the *lac* Operon

To fit the Setty *et al.* data (9), we used Eq. **25**, with s_1 corresponding to the logarithm of the IPTG concentration and s_2 corresponding to the logarithm of the cAMP concentration. As the base of the logarithm and a constant offset can be absorbed by the coefficients c_i, we chose to let $s_1 \in \{0, 1, \ldots, 5\}$ correspond to the six sample levels of IPTG and $s_2 \in \{0, 1, \ldots, 9\}$ correspond to the 10 sample levels of cAMP. We used a simplex search method (`fminsearch` in Matlab, Mathworks) to optimize the six parameters $c_0, c_1, \ldots, c_5$ so that the sum-squared error between Eq. **25** and the *lac* transcription data was minimized. We used multiple optimization runs and experimented with different initial conditions, but these factors seems to have little influence on the outcome of optimization. All or nearly all runs converged to essentially the same solution, which we therefore take to be close to optimal. The final parameters found were:

c_0	c_1	c_2	c_3	c_4	c_5
4.09	−1.88	0.15	−0.11	0.32	−0.06

which define the surface shown in Fig. 4*B*.

There is not a unique two state, bivariate Gaussian discrimination problem corresponding to these parameters (as described above). Of the many discrimination problem parameter sets consistent with the optimized c_i, we chose one by making the following assumptions:

$$(\boldsymbol{\sigma}^{\mathrm{I}})^{-1} = \begin{bmatrix} 0.4 & -c_3 \\ -c_3 & 0.3 \end{bmatrix} \qquad \textbf{[26]}$$

$$(\boldsymbol{\sigma}^{\mathrm{II}})^{-1} = (\boldsymbol{\sigma}^{\mathrm{I}})^{-1} + \begin{bmatrix} 2c_4 & 0 \\ 0 & 2c_5 \end{bmatrix} \qquad \textbf{[27]}$$

$$\boldsymbol{\mu}^{\mathrm{I}} = \begin{bmatrix} 1.5 \\ 3.5 \end{bmatrix} \qquad \textbf{[28]}$$

$$\boldsymbol{\mu}^{\mathrm{II}} = \left(\begin{bmatrix} -c_1 \\ -c_2 \end{bmatrix} + (\boldsymbol{\sigma}^{\mathrm{I}})^{-1} \boldsymbol{\mu}^{\mathrm{I}} \right) \boldsymbol{\sigma}^{\mathrm{II}} \qquad \textbf{[29]}$$

$$P(II) = (1 + \mathrm{e}^{z})^{-1} \qquad \textbf{[30]}$$

where

$$z \;=\; c_0 + \frac{1}{2}(\boldsymbol{\mu}^{\mathrm{I}})^T(\boldsymbol{\sigma}^{\mathrm{I}})^{-1}\boldsymbol{\mu}^{\mathrm{I}} - \frac{1}{2}(\boldsymbol{\mu}^{\mathrm{II}})^T(\boldsymbol{\sigma}^{\mathrm{II}})^{-1}\boldsymbol{\mu}^{\mathrm{II}} + \frac{1}{2}\log\left[\frac{\det(\boldsymbol{\sigma}^{\mathrm{I}})}{\det(\boldsymbol{\sigma}^{\mathrm{II}})}\right], \qquad [31]$$

which results in the distinct lognormal distributions in Fig. 4*C*. The five parameters unrecoverable from the posterior surface can be seen in our arbitrary choices for $\boldsymbol{\mu}^{\mathrm{I}}$, the diagonal elements of $(\boldsymbol{\sigma}^{\mathrm{I}})^{-1}$, and the off-diagonal elements (zero) added to $(\boldsymbol{\sigma}^{\mathrm{I}})^{-1}$ to make $(\boldsymbol{\sigma}^{\mathrm{II}})^{-1}$.

References

1. Monod J, Wyman J, Changeux JP (1965) *J Mol Biol* 12:88–118.
2. Shea MA, Ackers GK (1985) *J Mol Biol* 181:211–230.
3. Brown KS, Sethna JP (2003) *Phys Rev E* 68:21904.
4. Schwartz G (1978) *Ann Stat* 6:461–464.
5. Mackay DJC (2003) *Information theory, inference, and learning algorithms* (Cambridge University Press, New York, NY).
6. Stelling J, Gilles ED, Doyle FJ (2004) *Proc Natl Acad Sci USA* 101:13210–13215.
7. Gibson MA, Bruck J (2000) *J Phys Chem* 104:1876–1889.
8. Gillespie DT (1977) *J Phys Chem* 81:2340–2361.
9. Setty Y, Mayo AE, Surette MG, Alon U (2003) *Proc Natl Acad Sci USA* 100:7702–7707.