IOP Publishing Physical Biology

Phys. Biol. **4** (2007) 229–245 doi:10.1088/1478-3975/4/4/001

Cost–benefit theory and optimal design of gene regulation functions

Tomer Kalisky, Erez Dekel and Uri Alon

Department of Molecular Cell Biology and Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel

Received 31 May 2007
Accepted for publication 31 August 2007
Published 6 November 2007
Online at stacks.iop.org/PhysBio/4/229

Abstract

Cells respond to the environment by regulating the expression of genes according to environmental signals. The relation between the input signal level and the expression of the gene is called the gene regulation function. It is of interest to understand the shape of a gene regulation function in terms of the environment in which it has evolved and the basic constraints of biological systems. Here we address this by presenting a cost–benefit theory for gene regulation functions that takes into account temporally varying inputs in the environment and stochastic noise in the biological components. We apply this theory to the well-studied *lac* operon of *E. coli*. The present theory explains the shape of this regulation function in terms of temporal variation of the input signals, and of minimizing the deleterious effect of cell–cell variability in regulatory protein levels. We also apply the theory to understand the evolutionary tradeoffs in setting the number of regulatory proteins and for selection of feed-forward loops in genetic circuits. The present cost–benefit theory can be used to understand the shape of other gene regulatory functions in terms of environment and noise constraints.

M This article features online multimedia enhancements

(Some figures in this article are in colour only in the electronic version)

Introduction

Living organisms sense and respond to their environment in order to effectively utilize available resources and to survive stressful conditions. One way this is done at the cellular level is by employing circuits of interacting genes and proteins. Each such circuit acts as an input–output device, designed to be activated by a specific environmental signal (e.g. heat shock, osmotic stress or the presence of a sugar) and to elicit the required response (i.e. produce proteins that can deal with the stress or utilize the sugar). The response of genetic circuits is often not binary (being in either 'ON' or 'OFF' states); rather, it can be described by a continuous regulation function which relates the intensity of the input signal to the magnitude of output response.

It is of interest to understand the shape of the regulation function in terms of two major factors: first, the environment in which the organism evolves, where by environment we mean the distribution of the intensity of the input signals over time. The second factor that can affect the regulation function is an inherent constraint of biological material: circuits must function despite the stochastic nature of processes such as transcription and translation that result in cell–cell variability in the concentrations of the proteins that make up the circuits.

The effects of the environment on biological design can be treated within the framework of cost–benefit evolutionary analysis [1–4]. Here, the fitness of a circuit is viewed as the difference between the benefit and the cost to the fitness of the organism associated with the circuit. The cost of a design is defined as the reduction in fitness due to producing and maintaining the proteins that make it up. The benefit is the increase in fitness bestowed by the function of the circuit as it responds to its signal. Recently, cost and benefit were directly measured experimentally for a well-studied system in *E. coli*, the *lac* operon. Based on these measurements, cost–benefit theory was used to predict the optimal expression level of *lac* proteins in a given environment (defined by the concentration of the input signal, the sugar lactose). Evolutionary experiments then showed that cells evolve close to this predicted optimum within a few hundred generations

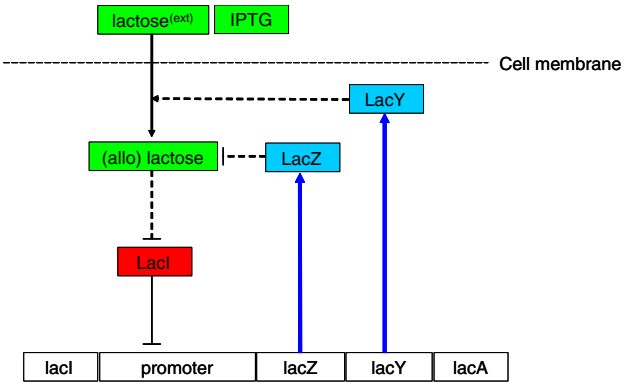

Figure 1. A diagram of the *lac* system in *E. coli*. Genes *lacZ* and *lacY*, which are responsible for lactose catabolism and transport, are transcriptionally repressed by the *lac* repressor LacI. When external lactose is introduced into the environment, it is initially transported by basal levels of LacY into the cell. Once lactose binds to LacI, its DNA binding is reduced, and repression is relieved.

[5]. These studies focused on constant environments in which the signal is present at constant levels throughout time. The cost–benefit theory was also extended to time-varying environments, where it was used to understand the selection of a common network motif called the feed-forward loop (FFL) [6] that can help to filter out brief signal pulses. Other works on evolution in time-varying environments have shown that population heterogeneity is beneficial when the environment changes much faster than the bacterial response time [7] or when the organism cannot sense effectively the transitions between different selective environments [8]. The effect of stochasticity within cost–benefit theory, however, has not been completely explored, nor has the shape of regulatory functions been studied in this light.

Here, we extend cost–benefit theory to understand the detailed shape of regulation functions. We use the *lac* operon as a model system to demonstrate how cost–benefit analysis can be used to study the effects of time-dependent pulses of signal in the environment and stochastic effects in the components of the circuit. This theory is shown to capture the measured shape of the regulation function of the *lac* operon. We further apply the present extended theory to understand the forces that determine the copy number of transcription factors, and to assign a new role to FFL circuits.

The rest of the introduction will discuss our model system, the *lac* operon. The *lac* operon in the bacterium *E. coli* is one of the classical examples of gene regulation mechanisms [9–11]. The *lac* operon consists of a set of genes that are activated in the presence of the sugar lactose in the environment. The regulation function of the *lac* operon defines the relation between the external lactose concentration and the steady-state production rate of the *lac* proteins [12, 13]. The experiments of [5] mentioned above suggest that evolutionary adaptation occurs rapidly, so that one may assume that the regulation function is optimally tuned by evolution to maximize fitness, i.e. it produces enough protein for effective lactose utilization while avoiding unnecessary protein production that will strain cell resources.

The *lac* operon encodes for three genes (*lacZ*, *lacY* and *lacA*) responsible for lactose catabolism and transport (figure 1). It is transcriptionally regulated by the repressor LacI and the activator cAMP-receptor-protein (CRP). The *lac* repressor, LacI, is a tetramer which is constitutively expressed at low concentrations (approximately 10 teramers/cell). In the absence of lactose, LacI binds strongly to the *lac* promoter, thus overlapping the RNA polymerase binding site and repressing transcription. However, when LacI binds to inducers such as allolactose or isopropyl-β-d-thiogalactoside (IPTG), its DNA binding is reduced and repression is relieved. The activator CRP binds the promoter only when bound to the inducer cyclic AMP (cAMP), which is a signaling molecule whose level depends on the starvation state of the cell. Glucose intake into the cells suppresses cAMP production and inactivates CRP. Thus, the *lac* operon is activated in the presence of lactose, but only if glucose is not simultaneously present. When active, the *lac* operon produces LacZ and LacY in large amounts (order of 10^4–10^5 copies/cell). The gene *lacY* encodes the β-galactoside permease LacY, a trans-membrane transport protein that pumps external lactose into the cell. *lacZ* encodes β-galactosidase (LacZ), an intracellular enzyme that cleaves the internal lactose into glucose and galactose. Some of the internal lactose is transformed by LacZ to allolactose, which then binds to LacI to induce production of LacZ and LacY. The precise role of *lacA* is currently not well understood, but it is thought to play a role in cellular detoxification [14, 15]. A simple mathematical model for the *lac* system, based on previous models by [16–18], is formulated in the appendix.

In the present study, the regulation function $\beta(L)$ of the *lac* system is defined as the production rate of *lac* proteins as a function of the external lactose concentration L, in the state of glucose starvation [12, 13]. The regulation function for wild-type *E. coli* strain MG1655, as measured by Dekel *et al* [5], is shown in figure 2(*a*). The regulation function can be approximated by a Hill function, i.e. $\beta(L) = \beta_{\mathrm{WT}} L^n/(L^n + K^n)$, where β_{WT} is the maximal production rate of *lac* proteins in wild-type (WT) *E. coli* at full induction, $K = 130\ \mu$M is the concentration for half-induction, and $n \approx 4$ is the Hill coefficient which determines the 'steepness' of the function. Saturation is reached at $L \approx 300\ \mu$M.

The shape of the regulation function depends on the regulation mechanism used by the cell (including effects of cooperativity [19], DNA looping [20], circuit architecture [21], etc), as well as the biochemical parameters of the circuit components (e.g., dissociation constants and reaction rates)—see the detailed model in the appendix. These parameters can be tuned by evolution in order to achieve an optimal regulation function shape for bacterial survival and growth. Furthermore, there are experimental indications that biochemical mechanisms are flexible enough to tailor the genetic regulation function to any of a wide range of possibilities [12].

In this paper, we wish to understand why the wild-type regulation function has evolved to its measured form, with specific slopes and saturation point. We assume that the regulation function has evolved to optimally suit the natural

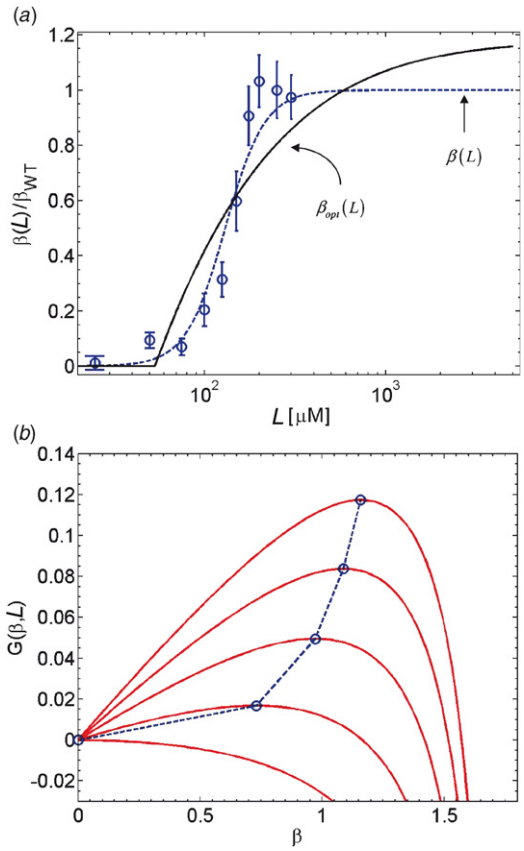

Figure 2. (*a*) The wild-type regulation function of the *lac* system $\beta(L)$ as measured by Dekel *et al* [5] (dashed line and experimental points) and the optimal expression for constant lactose environments $\beta_{\rm opt}(L)$ (solid line). Both functions were normalized to the maximal *lac* expression in wild-type *E. coli*: $\beta_{\rm WT}$. It can be seen that the two functions are different at high and intermediate lactose concentrations. The optimal expression was calculated according to experimentally supported cost–benefit analysis. The regulation function was measured using the ONPG colorimetric assay of β-galactosidase activity (see [5, 13] for details), and was fit to a Hill function $\beta(L) = \beta_{\rm WT} L^n/(L^n + K^n)$ with $K = 130$ μM and Hill coefficient $n = 4$ (dashed line). (*b*) The fitness function $G(\beta, L)$ for a constant lactose concentration L (solid lines). Shown are the curves for lactose concentrations of $L = 50$, 200, 500, 1200 and 5000 μM. Note that the fitness function is concave, with a negative second derivative. The optimal production rate $\beta_{\rm opt}(L)$ resides at the maximum of each curve (circles).

environment of *E. coli*. Hence, we study the effect of different environmental conditions on the optimal form of the regulation function.

The paper is organized as follows: we first demonstrate that optimality in constant environments cannot fully explain the detailed shape of the *lac* regulation function. We then explore how temporally varying environments, and the effect of cell–cell variation in regulatory protein concentration, can explain the observed shape in the context of cost–benefit theory. We show that (i) a pulsed lactose environment can explain the value of the wild-type regulation function at high lactose concentrations, but cannot explain its behavior at low to intermediate concentrations, (ii) introducing the evolutionary cost of noisy gene expression into the fitness function can explain the form of the regulation function for both high and intermediate regimes, independent of the temporal dynamics of the environment. Finally, we discuss the evolutionary factors that determine the concentration of transcription factors.

Results

Optimality in a constant environment cannot explain the shape of the lac regulation function

A cost–benefit theory for the *lac* operon under constant, defined conditions was established for exponentially growing *E. coli* MG1655 [5]. When considering constant lactose environments, with lactose concentration L, the *lac* expression level β leads to a growth burden (cost) due to production of the *lac* proteins, and a growth advantage (benefit) due to the metabolic activity of the *lac* proteins. The resulting fitness, defined as the relative change in growth rate, is:

$$G(\beta, L) = -\eta \frac{\beta}{1 - \beta/M} + \delta\beta \frac{L}{L + K_Y} \qquad (1)$$

where η, δ and M are experimentally measured constants, with $\eta/\delta = 0.12$ and $M = 1.8\beta_{\rm WT}$ ($\beta_{\rm WT}$ is the maximal production rate in wild-type *E. coli* at full induction). $K_Y = 0.4$ mM is the dissociation constant between lactose and LacY. η describes (to first order) the growth reduction per unit of *lac* protein expression. This is complemented by the constant M, which describes the maximal capacity for producing *lac* proteins: producing *lac* proteins at the rate approaching M would inhibit production of essential systems and significantly slow down the growth rate. δ is the relative growth advantage per *lac* expression unit at saturating lactose concentration. The cost of producing the *lac* proteins (first term) seems to originate from the limited resources of the cell (ribosomes, etc), and grows nonlinearly with the production rate β of the *lac* proteins (LacZ and LacY). The benefit (second term) is controlled by the rate of pumping lactose into the cell—see the appendix (equations (A.14)–(A.20)).

Given a lactose concentration L, the production rate which results in maximal fitness is given by finding the maximum of $G(\beta, L)$ (see figure 2(*a*)):

$$\beta_{\rm opt}(L) = M\left[1 - \left(\frac{\eta}{\delta} \cdot \frac{L + K_Y}{L}\right)^{1/2}\right] \qquad (2)$$

For lactose concentrations below a threshold $L_c = K_y(\delta/\eta - 1)^{-1} \approx 0.06$ mM, this optimal expression is negative, meaning that it is not beneficial to maintain the *lac* operon. In such environments, the ability to produce *lac* proteins is eventually lost [5].

As can be seen from figure 2(*a*), the predicted optimal expression level is close, to a first approximation, to the measured regulation function. The experimental accuracy in measuring this function is such, however, that significant differences can be noticed: the shape of the regulation function is significantly different from the optimal *lac* expression $\beta_{\rm opt}(L)$. The difference is prominent at high lactose concentrations ($L > 1000$ μM) where the optimal expression

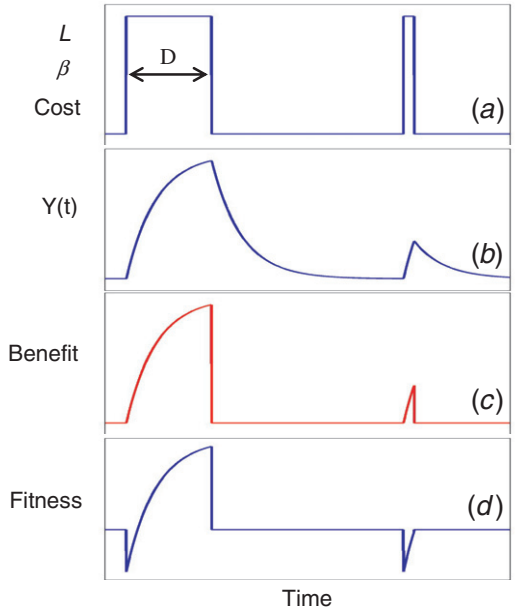

Figure 3. Cost and benefit for pulses of lactose. (*a*) Immediately at the beginning of the pulse the cell starts to produce the *lac* proteins (LacZ and LacY) at constant rate β. The cost is therefore constant throughout the duration of the pulse. (*b*) The amount of *lac* proteins increases on the time scale of cell generation. (*c*) The benefit depends on the level of *lac* proteins and external lactose. For short pulses the *lac* proteins do not accumulate in sufficient amounts before the pulse ends. In this case, the benefit does not surpass the cost and the resulting fitness (i.e. the difference between benefit and cost) is negative, as can be seen in (*d*).

is approximately 20% higher than the actual production rate. There is also a discrepancy at mid-range concentrations ($L < 400\ \mu$M) where the regulation function $\beta(L)$ is a steeper function than the predicted optimal function $\beta_{\rm opt}(L)$. In the following sections, we will investigate the effect of additional evolutionary factors influencing the design of the wild-type *lac* regulation function: variable environments and the inherent noise of the regulation mechanism.

Pulses of input signal in the environment may explain the shape of the regulation function at high lactose levels

Consider environments in which lactose is not present constantly over time, but rather appears in pulses of duration D and concentration L—see figure 3. This scenario may better represent the ecology of *E. coli* inside the mammalian host than constant lactose levels: lactose is thought to appear in the gut only rarely, that is, in rather isolated pulses that contain large amounts of sugar which are then rapidly depleted by the lactase enzymes of the host[1] [6, 22]. We will see in this section that the benefit of *lac* proteins is reduced if the pulse duration is shorter than the *response time* of the *lac* system, which is the time of roughly one cell generation required for *lac* proteins to accumulate to effective levels. Hence, the distribution of different pulse durations and amplitudes in the environment is a factor that may help to determine the optimal regulation function.

To form a cost–benefit theory, note that in the presence of lactose concentration L, the *lac* proteins are made at production rate β. When a pulse begins, the amount of *lac* proteins, Z, increases to its steady-state values according to [11] $Z(t) \sim \frac{\beta}{\alpha}(1-{\rm e}^{-\alpha t})$, where $\alpha = \ln 2/\tau_{\rm cell\text{–}cycle}$ is the dilution rate due to growth and division[2] (see the appendix). Therefore, the time-dependent fitness function during a pulse of duration D and amplitude L is given by:

$$G(\beta, L, t) = -\eta\frac{\beta}{1-\beta/M} + \delta\beta(1-{\rm e}^{-\alpha t})\frac{L}{L+K_Y}. \quad (3)$$

The benefit thus depends on the amount of *lac* proteins, which is zero at the beginning of the pulse and increases halfway to its maximal steady-state value only after one cell generation (roughly 30–60 min). The cost, on the other hand, is constant, because the bacterium starts producing *lac* proteins at rate β almost immediately (within 2–4 min) after pulse initiation (figure 3 and the appendix). Integrating over the pulse duration D gives the fitness function in a pulsing environment:

$$G_{\rm pulse}(\beta, L) = \int_{t=0}^{D} G(\beta, L, t)\,{\rm d}t = -\eta\frac{\beta}{1-\beta/M}D + \delta\beta\frac{L}{L+K_Y}\left[D + \frac{{\rm e}^{-\alpha D}-1}{\alpha}\right]. \quad (4)$$

Thus the optimal production rate in a pulsing environment is given by the value of β that maximizes equation (4) (figure 4(*a*))

$$\beta_{\rm opt}(L, D) = M\left[1 - \left(\frac{\eta}{\delta}\cdot\frac{L+K_Y}{L}\cdot A(D)\right)^{1/2}\right] \quad (5)$$

where

$$A(D) = \frac{1}{1+({\rm e}^{-\alpha D}-1)/\alpha D} \quad (6)$$

is the 'pulse factor' which decreases with pulse duration. Note that for very long pulses $A(D) \approx 1$, whereas for short pulses ($D \ll \tau_{\rm cell\text{–}generation}$): $A(D) \approx \frac{2}{\alpha D} \gg 1$.

As shown in figures 4(*a*) and (*b*), the optimal expression $\beta_{\rm opt}(L, D)$ depends on the duration of the pulse, and can be significantly lower than the optimal expression for constant lactose environments ($D \to \infty$). As the pulse duration D is shortened, higher concentrations of lactose are needed to generate the same optimal expression level.

It can be seen from figure 4(*b*) that for pulses shorter than $D_c = (2/\alpha)(\eta/\delta) \approx \tau_{\rm cell\text{–}generation}/3 \approx 10$ min the optimal expression is zero for any lactose concentration (see the appendix, equations (A.21)–(A.24)). This means that bacteria subject to pulses shorter than a critical pulse duration—roughly one third of a cell generation for the present case—gain no benefit from the *lac* operon [6]. The reason for this is that the response time of the system is one cell generation time. Hence, for shorter pulses the *lac* proteins will not reach a beneficial level before the pulse ends. This conforms to the view that

[1] Note that selection due to cost and benefit occurs only in the presence of lactose during the pulse.

[2] The concentrations of LacZ and LacY are proportional, with the proportionality factor determined by the different translation rates due to differences in the ribosome binding site of the two proteins.

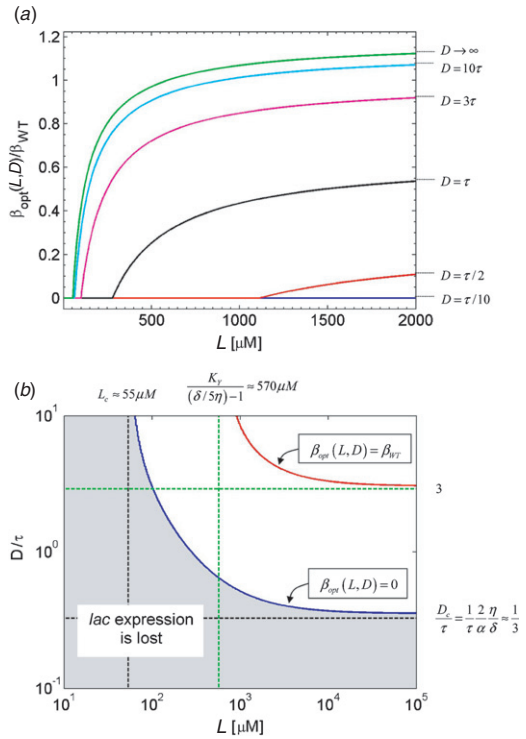

Figure 4. (*a*) The optimal expression level for pulsating environments. Shown are optimal production rates $\beta_{\rm opt}(L, D)$ for various amounts of lactose L and pulse durations D. It can be seen that as the pulse duration is shortened, larger amounts of lactose are required in order to maintain the same optimal expression level. τ represents the duration of one cell generation. (WT: wild-type). (*b*) A diagram showing the plane of pulse duration D versus pulse amplitude L (i.e., lactose concentration). The optimal production rate is zero for short pulses and low lactose concentrations. In this region the *lac* operon is not beneficial and will be lost in the course of evolution. Note that $\beta_{\rm opt}(L, D)$ is always zero if the pulse duration is less than approximately one third of a cell generation ($D \leqslant D_c = \frac{2}{\alpha}\frac{\eta}{\delta} \approx \frac{\tau}{3}$, where τ is the duration of one cell generation), independent of the lactose concentration. Additionally, it can be seen that the optimal expression $\beta_{\rm opt}(L, D)$ is equal to the maximal wild-type expression $\beta_{\rm WT}$ for a range of pulse durations and amplitudes. This may be a reason that at high lactose concentrations ($L > 1000\ \mu$M), the wild-type regulation function has lower values than the predicted optimal production rate in constant environments (see figure 2(*a*)).

genetic circuits in rapidly changing environments are effective only if their response time is equal to or faster than the rate of change of the environmental signals[3] (see also [7, 23]).

The above scenario may explain the value of the wild-type *lac* regulation function at high lactose concentrations ($L > 1000\ \mu$M): the wild-type *lac* protein expression is lower than the optimal expression for constant environments (figure 2(*a*)) due to the presence of lactose in pulses (figures 4(*a*) and (*b*)). However, the discrepancy at low to intermediate concentrations ($200\ \mu\text{M} < L < 400\ \mu\text{M}$) remains: the assumption of a pulsating environment cannot explain why the wild-type regulation function is higher than the predicted optimal expression in constant environments (equation (2)). We next turn to address this issue.

[3] Recent work has shown that rapidly switching between two selective environments can lead to heterogeneity [7]. However, here we assume that the time between successive pulses is very large relative to pulse duration.

Minimizing deleterious effect of noise may explain the regulation function shape at intermediate signal levels

In the previous section, we assumed that the environment of *E. coli* is characterized by short pulses of lactose. In this section we assume that lactose is present for long times (relative to the cell generation) and at various concentrations. We explore the possibility that the regulation function is optimally shaped to filter out cell–cell variability in the regulatory system.

The expression of proteins is noisy (stochastic), as has been well established experimentally (e.g. [24–36]). This noise results in variability between individual cells, which may have deleterious effects on fitness. Hence we study the fitness reduction due to noise, and show how the regulation function can be designed to minimize this reduction by adjusting its saturation points to the most probable lactose concentrations.

Throughout this section we assume that biochemical mechanisms in the cell are flexible enough to adjust the regulation function to any required shape [12]. Accordingly, we wish to derive the functional form of the regulation function of the *lac* system from first principles of evolutionary cost and benefit, without restricting ourselves to a specific biochemical implementation. Therefore, we assume a general scheme for a gene regulation system consisting of (i) an input signal L, (ii) a transcription factor R (activator/repressor), and (iii) an operon producing the required protein in response to the external signal according to the regulation function $F(L)$.

The fluctuations in the small number of repressor (LacI) molecules are a major source of noise in the lac system. The production of a protein by transcription and translation is a stochastic process [24–36]. The noise in protein p is usually quantified by the coefficient of variation (CV), η_p, defined as the standard deviation of protein abundance Δp divided by the mean $\langle p\rangle$. It was found that the CV follows a non-Poissonian relation [26, 28, 32]:

$$\eta_p^2 = \frac{1+b}{\langle p\rangle} \approx \frac{b}{\langle p\rangle}, \tag{7}$$

where $\langle p\rangle$ is the mean number of protein molecules per cell, and b is the 'burst size', i.e., the average number of proteins produced by a single mRNA molecule during its lifetime[4]. Thus, low abundance genes are usually the noisiest.

In the *lac* system, the noisy component is the *lac* repressor (LacI), which is expressed at very low concentrations (approximately 10 tetramers/cell). The large noise in LacI is transferred to its downstream genes[5] [31, 34]. This noise

[4] Recent experimental measurements [26, 32] suggest that b is on the order of 10 (bacteria) to 10^3 (yeast)—much larger than that expected in a Poissonian process (for which $b = 0$).

[5] *Intrinsic noise* is defined as variability caused by local stochastic events that occur during the process of gene expression. *Extrinsic noise* is defined as noise that has a global effect on the cellular environment. This includes cell–cell variability in global factors affecting gene expression (such as number of ribosomes or RNAp molecules) as well as low-copy transcription factors [31].

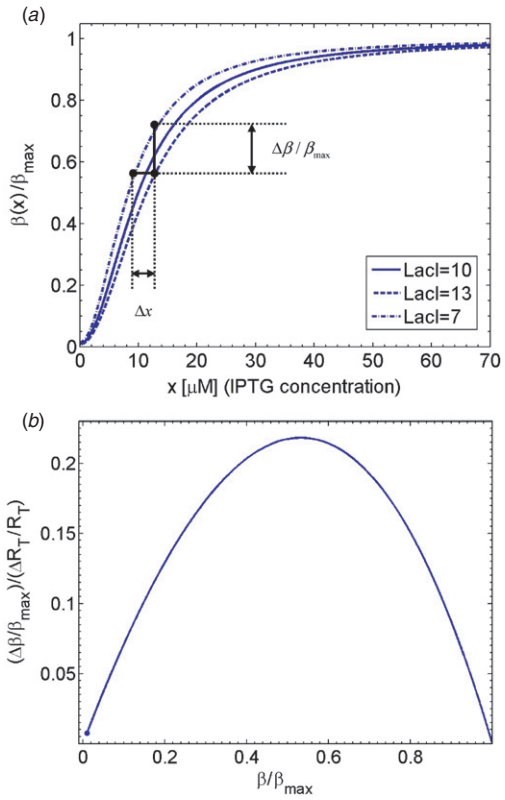

Figure 5. Extrinsic noise is maximal at intermediate lactose concentrations when the system is not fully activated or fully repressed, and is minimal at endpoints. (*a*) The regulation function of the *lac* operon when induced by the lactose analogue IPTG for different amounts of LacI molecules (see the appendix). The maximal noise is found at the point where the slope of the regulation function is maximal. (*b*) The noise versus the production rate. The noise is maximal for $\beta/\beta_{\max} \approx 1/2$. Similar results were obtained numerically for induction by lactose.

causes persistent cell–cell variability in *lac* protein levels (a type of extrinsic noise)[6] that can considerably affect the performance of the system.

Noise is maximal at intermediate lactose concentrations when the system is not fully induced or fully repressed, and is minimal at the saturation points. It was found experimentally [31, 34] that noise in *lac* protein levels displays a maximum at intermediate rates of transcription. The reason for this is that the regulation function of the *lac* system saturates at very high and very low concentrations of lactose, and is thus not susceptible to variability in LacI levels at the endpoints. At intermediate concentrations of lactose, the regulation function is most sharp and therefore sensitive to these variations (see figures 5(*a*) and (*b*)). Thus, setting the saturation points of the regulation function to match the most *frequent* lactose concentrations can best filter out extrinsic noise and increase the fitness of the organism.

In order to demonstrate this principle, we consider a standard model for the regulation of the *lac* system (as described in the appendix). The regulation function of the *lac* operon follows a Michaelis–Menten-like form:

$$\beta(L) = \beta_{\max} \cdot \frac{1}{1 + R_T g(L)/K_{\text{R–dna}}}, \tag{8}$$

where R_T is the total repressor concentration, $g(L)$ is a decreasing function of the external lactose L representing the fraction of active repressors (= DNA binding), and $K_{\text{R–dna}}$ is the repressor-DNA binding affinity. Assuming that the repressor concentration fluctuates by ΔR_T around its mean, one can show that the extrinsic noise $\Delta\beta$ in the production rate of *lac* proteins is given by (equations (A.25)–(A.30) in the appendix)

$$\frac{\Delta\beta}{\beta_{\max}} = \frac{\beta}{\beta_{\max}}\left(1 - \frac{\beta}{\beta_{\max}}\right) \cdot \frac{\Delta R_T}{R_T}. \tag{9}$$

The reason for this non-monotonic form (figure 5(*b*)) is that at high or low lactose concentrations, the *lac* repressor is either completely inactive (i.e. LacI cannot bind the DNA even at very large concentrations) or completely active (i.e. LacI binds the DNA at saturating concentration). Thus, the production rate at the endpoints does not depend significantly on the precise number of LacI molecules even in cases of large deviations from the mean (e.g. 10%). Thus, the extrinsic noise is zero at the saturation points ($\beta = 0$ and $\beta = \beta_{\max}$), and is maximal in the middle region where $\beta/\beta_{\max} = 1/2$.

The amplitude of the extrinsic noise is proportional to the derivative of the regulation function, and to the CV of the lac repressor. It can be noticed that the derivative of the regulation function is qualitatively similar to the magnitude of the noise in the regulation system: both vanish at the saturation points (zeros or full induction) and both display a maximum at intermediate activation levels. Actually, as shown in figure 5(*a*), a slight increase in repressor concentration is roughly equivalent to a slight decrease in external lactose concentration, since both reduce the gene expression. Thus the extrinsic noise behaves in a way similar to the derivative of the regulation function. The steeper the regulation function, the more vulnerable it is to variability in LacI.

For example, in a standard model of the *lac* system with a Hill-like regulation function $\beta(L)$, it can be shown that the magnitude of extrinsic noise as derived earlier (equation (9)) is also given by an expression that is proportional to the slope of the regulation function (see the appendix, equations (A.31)–(A.34)):

$$\Delta\beta \approx \frac{\mathrm{d}\beta}{\mathrm{d}L}\frac{\Delta R_T}{R_T} \cdot \frac{L}{n} \tag{10}$$

where n is the Hill coefficient. Hence, the standard deviation of the noise increases linearly with the derivative of the regulation function and the CV of the repressor concentration.

More generally, it is shown in the appendix that for *any* regulatory function $\beta(L)$ which is implemented using a transcription factor of concentration R_T, the noise is related to

[6] Note that although the intrinsic noise in the *production rate* of LacI has a short correlation time (on the order of burst duration [28, 33]), the persistence time of the variability in repressor *concentration* is dominated by the dilution rate, which is on the order of one cell generation. Thus, fast intrinsic fluctuations in LacI cause large cell–cell variability in *lac* proteins, which persist for long periods (order of one cell cycle) [36].

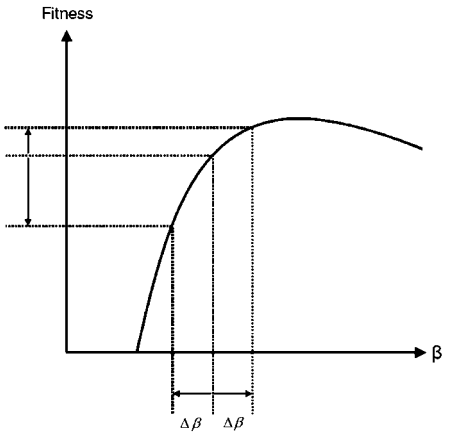

Figure 6. Extrinsic noise causes a reduction in the mean fitness. Shown schematically is the fitness function $G(\beta, L)$ for a certain lactose concentration L. The function is nonlinear, i.e. it is concave with a negative second derivative. It can be seen that symmetric fluctuations in the production rate β around the mean cause a larger fitness reduction than increase. The overall fitness reduction is proportional to $(\Delta\beta)^2$.

the steepness of the regulation function according to

$$\Delta\beta = \frac{\mathrm{d}\beta}{\mathrm{d}L}\frac{\Delta R_T}{R_T}\cdot B(L), \tag{11}$$

where $B(L)$ is some function of the external signal concentration L (e.g. lactose). The last relation is general for regulation systems consisting of a signal L, a transcription factor R_T, and a regulation function $\beta(L)$, and will be used in the following analysis.

Noise causes a reduction in the mean fitness, which is proportional to the noise variance. The fitness function of the *lac* operon for constant concentrations of lactose (equation (1)) is a nonlinear (concave) function of *lac* protein expression [5, 37] (see figure 2(*b*)). Due to the nonlinearity of the fitness function, symmetric fluctuations[7] around the mean expression level β result in a *non-symmetric* change in the total fitness of the population, as demonstrated in figure 6. Thus, in a population of cells, the magnitude of fitness reduction is larger than the magnitude of fitness increment, due to cell–cell variability. This results in an overall reduction in fitness, which is proportional to the curvature of the fitness function times the noise variance (see the appendix, equations (A.35)–(A.41)):

$$\langle G_{\text{population}}(\beta, L)\rangle = G(\beta, L) - \left|\frac{1}{2}\frac{\partial^2}{\partial\beta^2}G(\beta, L)\right|\cdot\langle\Delta\beta^2\rangle. \tag{12}$$

Substituting the previously derived expression for noise $\Delta\beta$ (equation (11)), and the expression for the non-noisy fitness function $G(\beta, L)$ (equation (1)), yields a cost–benefit expression with an additional term for the cost of the noise:

$$\langle G_{\text{population}}(\beta, L)\rangle = \left[-\eta\frac{\beta}{1-\beta/M} + \delta\beta\frac{L}{L+K_Y}\right] - \frac{C(\beta, L)}{R_T}\cdot\left(\frac{\mathrm{d}\beta}{\mathrm{d}L}\right)^2, \tag{13}$$

where $C(\beta, L)$ is a function of β and L (see the appendix). Note the factor $1/R_T$ results from the non-Poissonian cell–cell variability of LacI: $\langle\Delta R_T^2\rangle / R_T^2 = b/R_T$. One can see that the noise in the *lac* repressor adds an additional term to the fitness function, which is proportional to the square of the derivative of the regulation function $\beta(L)$. Thus, a steep regulation function $\beta(L)$ results in a reduction in the fitness which is proportional to the noise variance $\langle\Delta\beta^2\rangle$, or to the square of the slope of the regulation function. We next extend this to the case where the level of lactose is chosen from a distribution function.

The generalized fitness function includes the cost of noise, the cost of constitutively producing the repressor and the effect of the environment. Generally, given a probability distribution $P(L)$ for having lactose concentration L in the environment, and assuming that the regulation mechanism is realized by constitutively maintaining a repressor concentration of R_T, we suggest the following generalized fitness function:

$$J\{\beta(L)|P\} = -\eta_R\cdot R_T + \int_{L=0}^{\infty}\mathrm{d}L\cdot P(L)\cdot\left[-\eta\frac{\beta}{1-\beta/M} + \delta\beta\frac{L}{L+K_Y} - \frac{C(\beta, L)}{R_T}\cdot\left(\frac{\mathrm{d}\beta}{\mathrm{d}L}\right)^2\right], \tag{14}$$

for any given regulation function $\beta(L)$ (see the appendix, equations (A.35)–(A.41)).

The first term in equation (14) represents the cost of constitutively producing the *lac* repressor, whether lactose is present or not. The last term (inside the integral) represents the fitness reduction due to noise which results from the steepness of the regulation function $\beta(L)$. The cost of noise is reduced as higher repressor concentrations (R_T) are maintained. The two remaining terms (first and second terms in the integral) represent the cost and benefit of producing the *lac* proteins at rate $\beta(L)$ at external lactose concentration L [5] as formulated in equation (1). An optimal regulation function $\beta(L)$ that maximizes the generalized fitness function (equation (14)) will tend to follow the general trend of $\beta_{\text{opt}}(L)$ (equation (2)), but will also tend to saturate at the most probable lactose concentrations in order to minimize the deleterious effects of noise.

In the following sections, we will show how optimization of the generalized fitness function can lead to the observed shape of the *lac* regulation function, as well as the small concentration of the repressor.

Different environments favor different optimal regulation functions. Given a lactose probability distribution $P(L)$ and a given repressor concentration R_T, it is possible to find the function $\beta(L)$ that maximizes the generalized fitness function (equation (14)) using a variational approach (solved numerically by gradient based methods, see the appendix).

[7] Note that in general, noise in gene expression is not symmetric around the mean. Our assumption of symmetry is more of a 'first-order' approximation, and is valid when the noise distribution is not very skewed.

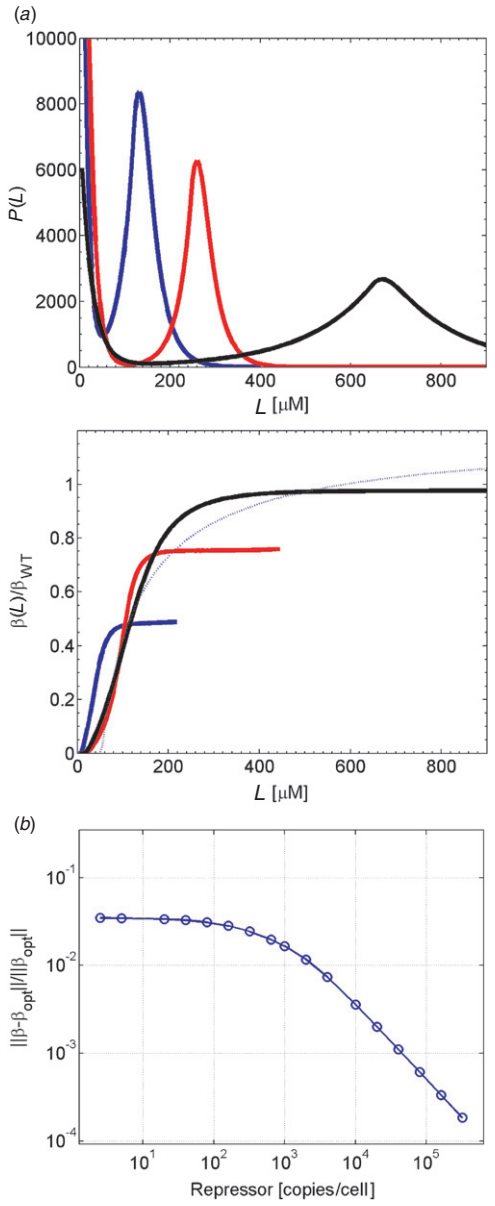

Figure 7. (*a*) Different environments favor different regulation functions. Shown is the optimal regulation function $\beta(L)$ as calculated from the generalized fitness function by the gradient-based method for various environments $P(L)$. It can be seen that the saturation points of the regulation function $\beta(L)$ correspond to the lactose concentrations where lactose is most abundant (i.e., to the maxima of $P(L)$). This analysis suggests that the wild-type regulation function is optimally suited for environments in which lactose concentrations of roughly $L = 650$ μM and $L = 0$ are most abundant. (*b*) The relative deviation of $\beta(L)$ from $\beta_{\rm opt}(L)$ as a function of repressor level. The distance metric between functions is defined as $\|\beta - \beta_{\rm opt}\| = (\int_{L=0}^{\infty} [\beta(L) - \beta_{\rm opt}(L)]^2 \, {\rm d}L)^{1/2}$. It is seen from the figure that large transcription-factor concentrations enable sharp regulation functions to be maintained (i.e., closer to $\beta_{\rm opt}(L)$), because the extrinsic noise is small. $\beta(L)$ was calculated from the generalized fitness function for an environment $P(L)$ with maxima at $L = 0$ and $L = 650$ μM.

In figure 7(*a*) we show different regulation functions that are optimally suited for various environments. We explored environments where a bacterium is exposed to mainly two states ('working points') of high and low lactose concentration. Each such environment is described by a bimodal form for the distribution $P(L)$. It can be seen that the optimal regulation function saturates at points of maximum lactose probability. When saturated, ${\rm d}\beta/{\rm d}z$ is small, resulting in effective noise repression at the most common environmental situations. The present theory suggests that the wild-type regulation function is optimally suited for an environment in which lactose concentrations of roughly $L = 650$ μM and $L = 0$ are most abundant [5] (figure 7(*a*)).

In figure 7(*b*) we show the relative deviation from $\beta_{\rm opt}(L)$ due to the cost of noise, as a function of the number of repressor molecules. It can be seen that for large concentrations of the transcription factor, steeper functions can be maintained because there is less extrinsic noise. However, as will be shown next, constitutively producing large amounts of the *lac* repressor (LacI) comes at a significant cost.

So far we discussed bi-modal lactose distributions. We also tested unimodal distributions that peak around a typical lactose concentration L_0. We find that unimodal distributions of lactose $P(L)$ tend to show a regulation function with a step: a function that follows $\beta_{\rm opt}(L)$ at values of L far from the peak L_0, but shows a flattened step near the peak concentration. This flattened step reduces the effects of variability at the most common lactose concentration around L_0. Since the observed regulation function has no such step, one may conclude that $P(L)$ in the environment is not unimodal.

Cost–benefit analysis of constitutively expressed regulatory proteins

In order to implement the *lac* regulation function, constitutive expression of repressor molecules (LacI) is maintained by the cell at all times. The bacterium maintains a concentration of about 10 LacI tetramers per cell. We now ask: what determines this concentration of LacI? We suggest that this may be related to the sharpness of the required regulation function $\beta(L)$ and to the accompanying noise.

The generalized fitness function with respect to LacI concentration (R_T) can be written as

$$J(R_T) = -\eta_R \cdot R_T - \frac{N}{R_T}, \tag{15}$$

where $N = \int_{L=0}^{\infty} {\rm d}L \cdot P(L) \cdot C(\beta, L) \cdot \left(\frac{{\rm d}\beta}{{\rm d}L}\right)^2$ is the cost of noise (figure 8). There are thus two opposing evolutionary forces tuning the number of repressor molecules: (i) constitutively producing the repressor—even when not needed—confers a continuous burden on cell resources (first term), (ii) maintaining a large repressor concentration reduces cell–cell variability and reduces the deleterious effect of noise on fitness (second term).

Optimizing with respect to R_T yields an optimal concentration that increases with the cost of noise:

$$R_T^{(\rm opt)} = \left(\frac{N}{\eta_R}\right)^{1/2}. \tag{16}$$

Thus, noise-sensitive systems, in which the noise is not filtered out effectively by the regulation function (large N), require a

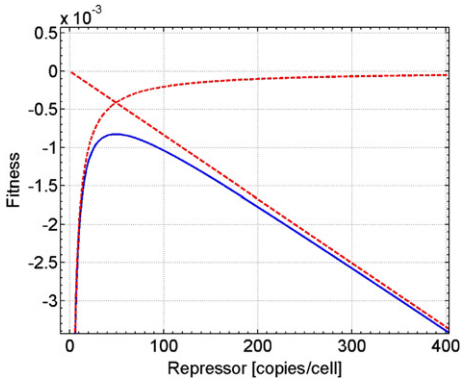

Figure 8. Cost–benefit analysis for *lac* repressor concentration (R_T). Shown is the generalized fitness function for various amounts of LacI. There are two opposing evolutionary forces: (i) the cost of constitutively producing LacI ($-\eta_R \cdot R_T$) forces the cell to produce the *lac* repressor in small amounts, (ii) the benefit of large amounts of LacI to noise reduction ($-N/R_T$) compels the cell to maintain many LacI molecules. We suggest that in the *lac* system the result is a small repressor concentration, and a regulation function which is optimally tuned to cope with extrinsic noise.

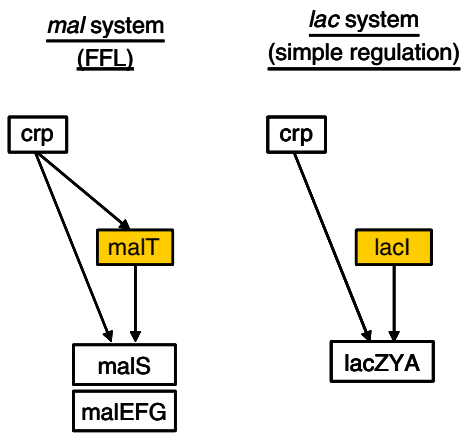

Figure 9. The feed-forward loop can allow higher concentrations of regulator, reducing cell–cell variability. Shown is a sketch of the lactose (*lac*) and the maltose (*mal*) utilization systems. The *lac* system is regulated by two transcription factors: LacI and the global regulator CRP, which is activated upon glucose starvation. The cost of constitutively expressing LacI forces the cell to produce it in small amounts, resulting in large cell–cell fluctuations. As opposed to this, the *mal* system is built as a feed-forward loop (FFL). This architecture enables production of large amounts of the transcription factor (MalT), but only when needed. Thus, the extrinsic noise can be smaller in the FFL architecture. The disadvantage of the FFL is that it operates at a delay on the order of cell generation (the time it takes for MalT to accumulate to its operational level), and it is possible that this constraint is unacceptable for the *lac* system.

large number of transcription factor molecules to be constantly produced in order to avoid high noise levels. However, if the regulation function filters out the noise effectively (the parameter N is small) a low transcription factor concentration is sufficient.

This result might be extended to other regulators. One should then consider the number of regulators per gene regulated. Each regulated gene will contribute its own 'noise cost'. Note that constitutively expressed regulators with many targets (such as CRP) tend to have higher copy numbers, but roughly similar copy numbers *per regulated gene* compared to LacI. A possible interpretation of this might be that regulation functions of bacterial genes are designed to minimize the cost of noise N to a common tolerated level, thus requiring constitutive production of transcription factors at roughly the same small concentration per regulated gene.

A role for feed-forward- loops in minimizing cell–cell variability

So far we have considered the *lac* operon which is a simple gene regulation circuit. We now briefly apply the present theory to circuits with slightly more complex architecture, feed-forward loops. Feed-forward loops (FFLs) are common network motifs in transcription regulation networks [6, 11, 21, 38–42]. An FFL is composed of two genes X and Y which regulate a common target gene Z. In the FFL there is an additional arrow of regulation between X and Y (see the sketch in figure 9). FFLs are thought to function as sign-sensitive filters that filter out short activation/inactivation signals in gene X [6]. Here we propose an additional role: the FFL enables a regulation system to produce large amounts of the transcription factor Y, but only when required. This enables a considerable reduction in noise, at the cost of a delayed response [6].

For example, in the *lac* system CRP and LacI both regulate the *lac* operon. The cost of constitutively producing LacI may drive selection toward production of very low concentrations of the *lac* repressor (10 tetramers/cell [10]), resulting in strong stochastic noise. However, not all regulators are expressed in such small amounts. An example occurs in the *maltose* regulon: CRP and the transcription factor *malT* both regulate the maltose uptake and metabolism genes (*malS*, *malEFG*, *malK*, etc) [43, 44]. However, in the *maltose* system, CRP also regulates *mal*T, thus forming an FFL. Hence, MalT protein is significantly produced only in times of glucose starvation (which activates CRP). Since MalT in this system is not constitutively expressed, the cell does not need to bear its cost under all conditions. In fact, no cost of MalT production occurs unless CRP is active. Hence, the cell can 'afford' to produce MalT in large amounts—up to a few hundred of monomers per cell [43]. As a result, the stochastic effects in the maltose system are expected to be smaller than in the *lac* system, when these systems are induced.

Similar behavior can be observed in FFLs of the *arabinose* and the *galactose* utilization systems, where the transcription factors *araC* [45–47] and *galS* [40, 48, 49] are also regulated by CRP, and appear in large concentrations (∼100 copies/cell) when activated by CRP, while the transcription factor *galR*, which is constitutively expressed, appears in small concentrations (∼10 copies/cell). Thus, a beneficial effect of these FFLs is to allow enhanced production of regulators, thereby helping to reduce stochastic effects.

Not every system, however, can employ an FFL for this purpose. Dekel *et al* [6] analyzed conditions for selection of FFLs as compared to simple regulation, showing that FFLs

are selected in environments with both short and long input pulses, and where the cost–benefit ratio of the system is not too high. Future work can study FFL selection in detail, using the present approach, including both noise and environmental factors.

Conclusion and outlook

This paper presented an extended cost–benefit theory that allows prediction of the optimal gene regulation function in terms of the distribution of input signal in the environment and in terms of minimizing the adverse effect of noise due to variations in regulatory proteins. A generalized fitness function that takes these effects into account was presented. This theory appears to help understand the detailed shape of the regulation function of the *lac* operon. The theory predicts that high lactose levels appear in the environment of *E. coli* only in short pulses, resulting in lower values for the regulation function than expected in constant environments. These lower values are due to the reduced benefit of the *lac* proteins during brief pulses. The steep shape of the regulation function at intermediate lactose levels is suggested to optimally minimize the effects of noise in an environment with a bimodal distribution of lactose concentrations. Finally, we suggest an explanation for the small concentration of the *lac* repressor in terms of the balance between noise cost and regulatory system cost[8].

At present, much more is known about molecular regulatory mechanisms than about the ecology in which they evolved. Even for the *lac* system, knowledge is limited as to the lactose pulse duration, frequency and concentrations found naturally [22, 50]. The present approach therefore contributes to the effort to infer facts about the natural environment of an organism from the architecture of its genetic circuits (an inverse problem that may be called 'inverse ecology' [6]). The assumption that bacteria can optimize their regulation functions rapidly (within hundreds of generations as suggested by experiments) allows one to relate the observed functions to those predicted to be optimal in different environments. In the present case, for example, one can suggest that the signal is most commonly found in the environment at concentrations where the regulation function is least prone to noise, i.e., the saturation points of the regulation function. Furthermore, we suggest that regulation functions whose values are lower than those predicted to be optimal in a constant environment, indicate the possibility that the external signal appears in pulses. The accurate experimental measurement of cost and benefit parameters [5] is essential in order to solve this inverse problem.

[8] Cost–benefit analysis could also be applied to study the role of the third gene in the *lac* operon: LacA (galactoside acetyltransferase). It has been proposed [14, 15] that LacA plays a role in cellular detoxification by eliminating non-metabolizable lactose analogs, which would otherwise accumulate and have a detrimental effect on cell growth. This is done by transferring acetyl groups to these non-metabolizable compounds thereby removing them from the cell and inhibiting their retransportation by LacY. Future work could generalize our analysis to include the cost of producing LacA versus the benefit of removing interfering compounds.

It would be interesting to apply the present theory, in conjunction with accurate experimental measurements, to try to understand the detailed shape of input functions of other genes. Experiments using fluorescent reporter strains can help to achieve the desired experimental accuracy [51]. Evolutionary laboratory lab experiments [3, 5] can help to test the present theory, by measuring how regulatory functions evolve in defined environments (such as environments with pulses of input signals). New experimental techniques that enable long-term monitoring of cells in well-controlled environments [52], may enhance the feasibility of such evolution experiments. We believe that the approach presented in this paper will be of use in such studies, as well as in understanding how the natural design of gene regulatory functions are shaped by the environment and noise constraints of the cell.

Acknowledgments

We wish to thank Robert Schleif for useful advice. We thank Shalev Itzkovitz, Avi Mayo, Erez Oxman, Guy Shinar, Anat Bren, Alon Zaslaver, Shai Kaplan, Ariel Cohen, Alex Sigal, Nadav Kashtan, Merav Parter, and all members of our lab for helpful discussions and feedback. We thank the Kahn Family Foundation for support. Tomer Kalisky was supported by the Israeli Center for Complexity Science.

Appendix

A.1. A simple model for the lac operon

Throughout this paper, we use the following equations for modeling the *lac* system [5, 16–18] (see the sketch in figure A3 and table A1).

External lactose ($L_{(e)}$) is pumped into the cell by LacY (Y) and utilized by LacZ (Z). The dynamics of internal lactose concentration (L) is given by:

$$\frac{\mathrm{d}L}{\mathrm{d}t} = v_y[L_{(e)}Y] - \bar{v}_y[LY] - (v_{z1} + v_{z2})[LZ] - \tau^{-1}_{\text{dilution}} L$$
$$\Rightarrow v_y[L_{(e)}Y] = (v_{z1} + v_{z2})[LZ]. \quad \text{(A.1)}$$

Some of the internal lactose utilized by LacZ, and some is transformed into allolactose (A):

$$\frac{\mathrm{d}A}{\mathrm{d}t} = v_{z2}[LZ] - v_{z3}[AZ] - \tau^{-1}_{\text{dilution}} A$$
$$\Rightarrow v_{z2}[LZ] = v_{z3}[AZ]. \quad \text{(A.2)}$$

Allolactose binds LacI (R) and deactivates it, thus releasing it from the DNA:

$$\frac{R}{R_T} = \frac{1}{1 + (A/K_{A-R})^n} \quad (n \approx 2\ [17]). \quad \text{(A.3)}$$

The repression is relieved, and LacZ and LacY are produced:

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \beta_z \cdot \frac{1}{1 + \left(\frac{R}{K_{\text{R-dna}}}\right)} - \tau^{-1}_{\text{dilution}} \cdot Z \quad \text{(A.4)}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \beta_y \cdot \frac{1}{1 + \left(\frac{R}{K_{\text{R-dna}}}\right)} - \tau^{-1}_{\text{dilution}} \cdot Y. \quad \text{(A.5)}$$

Note the separation of timescales:

- Equations (A.1) and (A.2) describe reactions that reach steady state within short times (fractions of a minute)

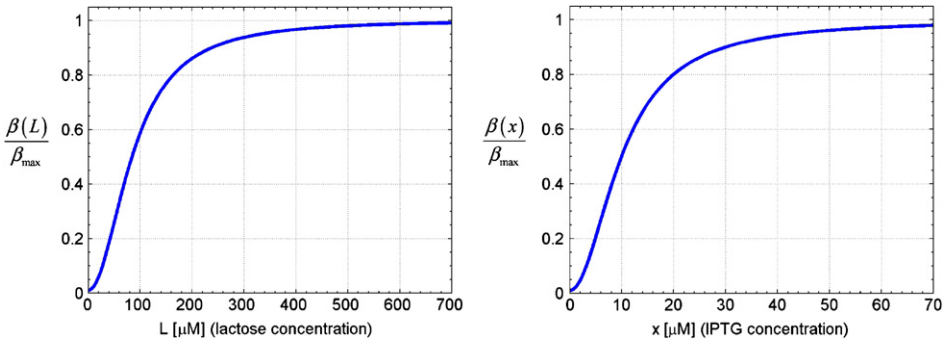

Figure A1. The regulation function of the *lac* operon with lactose and IPTG, as calculated by numerical integration of equations (A.1)–(A.8).

Table A1. Variables and constants used for modeling the *lac* operon.

Parameter	Definition	Value
$L_{(e)}$	External lactose concentration	
X	External IPTG or TMG	
L	Internal lactose concentration	
A	Allolactose concentration	
R	Active (=DNA binding) LacI concentration	
Z	LacZ concentration	
Y	LacY concentration	
$\beta_{y,z}(L)$	The regulation *function* expressing the production rate of LacZ (or LacY) as a function of external lactose (or IPTG) concentration L	
R_T	Total amount of repressor (constant)	0.01 μM (10 molecules) [10]
Z_T	Maximal amount of LacZ	50 μM ($\sim 10^4$ molecules) [10]
Y_T	maximal amount of LacY	50 μM ($\sim 10^4$ molecules) [10]
$\tau_{\text{cell–generation}}$	Cell generation	30 minutes [17]
α, $\tau^{-1}_{\text{dilution}}$	Dilution rate	$\frac{\ln 2}{\tau_{\text{cell–generation}}} = 2.26 \times 10^{-2}$ min^{-1} $\approx$ 1/(50 min)
β_z (β_{WT}, $\beta_{\max}$)	Maximal production rate of LacZ	$\sim 10^2$ molecules min^{-1} [16]
β_y (β_{WT}, $\beta_{\max}$)	Maximal production rate of LacY	$\sim 10^2$ molecules min^{-1} [16]
$K_{\text{TMG}-y}$	Affinity of TMG and LacY	700 μM [53]
$K_{L(e)-y}$	Affinity of external lactose and LacY	400 μM [5]
K_{L-y}	Affinity of (internal) lactose and LacY	1.8 mM [5]
K_{L-Z}	Affinity of (internal) lactose and LacZ	1.4 mM [5]
K_{A-Z}	Affinity of Allolactose and LacZ	1.9 mM [17]
K_{A-R}	Affinity of Allolactose and LacI	6 μM [17]
$K_{\text{IPTG–R}}$	Affinity of IPTG and LacI	Assumed to be 1 μM
$K_{\text{TMG–R}}$	Affinity of TMG and LacI	Assumed to be 6.3 μM
$K_{\text{R–dna}}$	Affinity of LacI and the DNA	10^{-10}–10^{-11} M [53]
v_y	Velocity of pumping by LacY	3000 molecules min^{-1} [17]
$\tilde{v}_y$	Velocity of efflux by LacY	$v_y/100$ [5]
v_{z1}	Velocity of internal lactose hydrolysis by LacZ	$v_{z1} = 0.9v_y$ [5]
v_{z2}	Velocity of conversion of lactose to allolactose by LacZ	20 000 molecules min^{-1} [17]
v_{z3}	Velocity of allolactose hydrolysis by LacZ	20 000 molecules min^{-1} [17]

relative to the cell cycle due to the fast utilization rates of internal lactose and allolactose by LacZ $\left((v_{z1}+v_{z2})[LZ] \gg \tau^{-1}_{\text{dilution}}L;\, v_{z3}[AZ] \gg \tau^{-1}_{\text{dilution}}A\right)$. Hence a steady-state metabolic flow can be assumed above this timescale.

- Equation (A.3) also describes a fast reaction (protein-ligand binding and protein activation), reaching steady state after roughly 1 ms.
- Equations (A.4) and (A.5) describe the production of LacZ and LacY through transcription and translation. They are valid for timescales above 5 min, and will reach halfway to steady state only after a cell generation (Binding of LacI to the DNA will reach steady state after 1s, mRNA lifetime is roughly 2 min, and translation will reach steady state after approximately 5 min. Thus the production rate reaches steady state after 5 min).

Note also that in equation (A.1) we have neglected the efflux component.

In order to be activated, the *lac* operon needs to maintain a small basal level of LacY pumps in order to detect small amounts of lactose. Figure A1 shows the response function for different amounts of lactose, as calculated by numerical integration of equations (A.1)–(A.5).

A.2. Induction by IPTG

Isopropyl-β-d-thiogalactoside (IPTG) is a lactose analog that can bind to LacI, but enters the cell even in the absence of LacY and is not utilized by LacZ. The model equations for IPTG are

$$\frac{R}{R_T} = \frac{1}{1+(x/K_{\mathrm{IPTG-R}})^2} \tag{A.6}$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \beta_z \cdot \frac{1}{1+\left(\frac{R}{K_{\mathrm{R-dna}}}\right)} - \tau_{\mathrm{dilution}}^{-1} \cdot Z \tag{A.7}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \beta_y \cdot \frac{1}{1+\left(\frac{R}{K_{\mathrm{R-dna}}}\right)} - \tau_{\mathrm{dilution}}^{-1} \cdot Y, \tag{A.8}$$

where x represents the IPTG concentration. The production rate is (reaching steady state after about 5 min):

$$\beta_{y,z}(x) \equiv \beta_{y,z} \cdot \frac{1}{1+\frac{R_T}{K_{\mathrm{R-dna}}}\frac{1}{1+(x/K_{\mathrm{IPTG-R}})^2}}. \tag{A.9}$$

The response function for IPTG is simulated in figure A1.

A.3. Induction by TMG

Thio-methyl-galactoside (TMG) is a lactose analog that is pumped into the cell by LacY, and also binds to LacI, but is not utilized by LacZ. It was shown by Ozbudak *et al* [53] that TMG causes a strong positive feedback loop which results in bistable behavior. For completeness, we show that the present simple model captures this bistability. The model equations for TMG are

$$\frac{\mathrm{d}x}{\mathrm{d}t} = v_y[x_{(e)}Y] - \tau_{\mathrm{dilution}}^{-1}x \tag{A.10}$$

$$\frac{R}{R_T} = \frac{1}{1+(x/K_{\mathrm{TMG-R}})^2} \tag{A.11}$$

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \beta_z \cdot \frac{1}{1+\left(\frac{R}{K_{\mathrm{R-dna}}}\right)} - \tau_{\mathrm{dilution}}^{-1} \cdot Z \tag{A.12}$$

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \beta_y \cdot \frac{1}{1+\left(\frac{R}{K_{\mathrm{R-dna}}}\right)} - \tau_{\mathrm{dilution}}^{-1} \cdot Y, \tag{A.13}$$

where x represents the TMG concentration. Note that the dilution rate of internal TMG in equation (A.10) cannot be neglected this time, because TMG is not degraded by LacZ. The response function for TMG (i.e., the concentration of LacZ and LacY at steady state), as calculated by numerical integration of equations (A.10)–(A.13), is shown in figure A2. It can be seen that for a range of TMG concentrations, initial [LacY] = 0 results in no response, while initial [LacY] > 0 results in maximal response (provided that sufficient time has elapsed after induction). Thus, due to the strong positive feedback mechanism, different initial concentrations of LacY result in different steady states.

A.4. Optimization in a pulsating environment

Assume that at time point $t = 0$ a concentration L of lactose is introduced into the external environment of the cells. In roughly 5 min the production rates of LacZ and LacY will reach their steady-state values $\beta_{y,z}$, and the system can be described by equations (A.1)–(A.5). The concentrations of

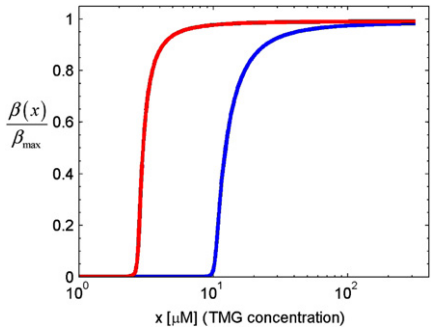

Figure A2. The bistable response function of the *lac* operon for TMG, as calculated by numerical integration of equations (A.10)–(A.13) under different initial conditions for LacY concentration. The blue curve corresponds to initial concentration [LacY] = 0 while the red curve corresponds to initial concentration [LacY] = 0.1 μM. Results are for 200 min after induction.

LacZ and LacY increase according to equations (A.4) and (A.5) as (note that $\tau_{\mathrm{dilution}}^{-1} = \alpha$):

$$Z(t) = \frac{\beta_z}{\alpha}(1-\mathrm{e}^{-\alpha t}), \qquad Y(t) = \frac{\beta_y}{\alpha}(1-\mathrm{e}^{-\alpha t}). \tag{A.14}$$

Thus, the cost, which is a nonlinear function of the production rate β [5], reaches its steady state after roughly 5 min:

$$\mathrm{cost} = -\eta\frac{\beta}{1-\beta/M}. \tag{A.15}$$

The cost parameter η was measured experimentally [5]—see table A2. The benefit is a function of the lactose and allolactose utilization rate by LacZ, which is limited by the amount of available pumps (LacY)—see figure A3:

$$\begin{aligned}\mathrm{benefit} &\sim v_{z1}[ZL] + v_{Z3}[ZA] = (v_{z1}+v_{z2})\,[ZL] \\ &= v_y[YL_{(e)}] = v_y Y\frac{L_{(e)}}{L_{(e)}+K_{L(e)-Y}}.\end{aligned} \tag{A.16}$$

Substituting $Y(t)$ gives

$$\mathrm{benefit} \sim v_y \cdot \frac{\beta_y}{\alpha}(1-\mathrm{e}^{-\alpha t}) \cdot \frac{L_{(e)}}{L_{(e)}+K_{L(e)-Y}}, \tag{A.17}$$

which approaches its steady state only after one cell generation (on the order of 50 min). Hence, we get the time-dependent fitness function:

$$G(\beta, L, t) = -\eta\frac{\beta}{1-\beta/M} + \delta\beta(1-\mathrm{e}^{-\alpha t})\frac{L}{L+K_Y}. \tag{A.18}$$

Integrating over one pulse duration D gives the fitness function for one pulse:

$$\begin{aligned}G_{\mathrm{pulse}}(\beta, L) &= \int_{t=0}^{D} G(\beta, L, t)\,\mathrm{d}t = -\eta\frac{\beta}{1-\beta/M}D \\ &+ \delta\beta\frac{L}{L+K_Y}\left[D + \frac{\mathrm{e}^{-\alpha D}-1}{\alpha}\right].\end{aligned} \tag{A.19}$$

The optimal production rate in a pulsating environment is found by maximizing $G_{\mathrm{pulse}}(\beta, L)$ with respect to β:

$$\beta_{\mathrm{opt}}(L, D) = M\left[1-\left(\frac{\eta}{\delta}\cdot\frac{L+K_Y}{L}\cdot A(D)\right)^{1/2}\right], \tag{A.20}$$

where $A(D) = \frac{1}{1+(\mathrm{e}^{-\alpha D}-1)/\alpha D}$ is the pulse factor. For very long pulses $A(D) \approx 1$, whereas for short pulses ($D \ll \tau_{\mathrm{cell-generation}}$): $A(D) \approx \frac{2}{\alpha D} \gg 1$.

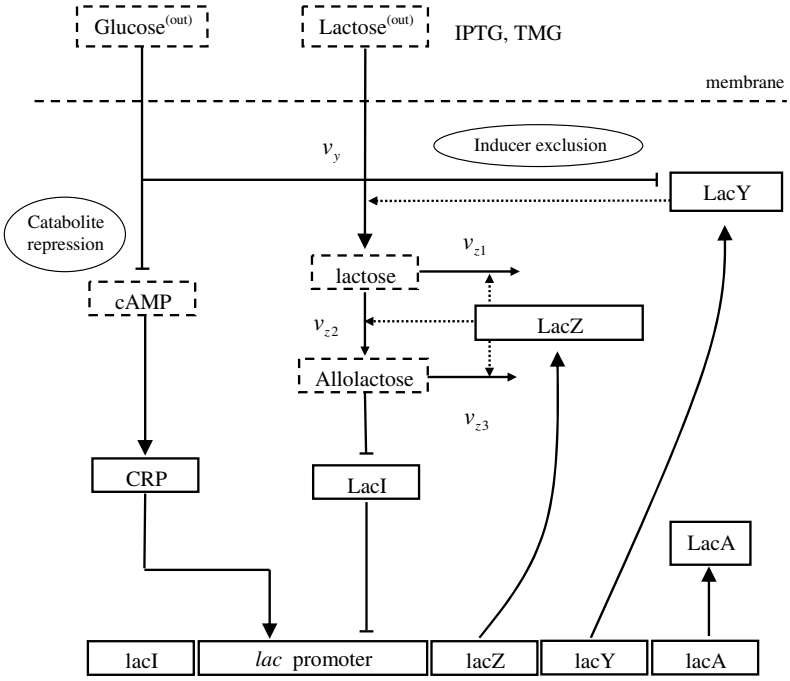

Figure A3. A detailed sketch of the *lac* system.

Table A2. Constants used for cost-benefit analysis of the *lac* operon as measured by Dekel *et al* [5].

Parameter	Definition	Value
η	Coefficient of cost for producing LacZ and LacY	$0.02/\beta_z$
η_R	Coefficient of cost for producing LacI	We assume that $\eta_R \approx \eta$
δ	Coefficient of benefit of producing LacZ and LacY	$0.17/\beta_z$
M	Upper boundary on production rate of LacZ	$1.8\,\beta_z$

Note that we have assumed that the dilution rate α, which depends on the growth rate during the pulse, does not depend significantly on the pulse amplitude L or duration D. This is justified if the underlying medium is rich enough to enable growth even without lactose (e.g. medium supplemented with glycerol [5]).

A.5. For very short pulses the lac system is not beneficial

We will now find the conditions for which $\beta_{\rm opt}(L, D) = 0$. This happens when

$$\frac{\eta}{\delta} \cdot \frac{L + K_Y}{L} \cdot A(D) = 1. \tag{A.21}$$

For short pulses ($D \ll \tau_{\rm cell-generation}$) we substitute $A(D) \approx 2/\alpha D$:

$$\frac{\eta}{\delta} \cdot \frac{L + K_Y}{L} \cdot \frac{2}{\alpha D} = 1 \tag{A.22}$$

which gives the curve [6]

$$D = \frac{2}{\alpha} \cdot \frac{\eta}{\delta} \cdot \frac{L + K_Y}{L} \tag{A.23}$$

in the D–L plane. The asymptote is found by substituting $L \to \infty$ which gives (see figure 4(*b*) in the main text):

$$D_c = \frac{2}{\alpha} \cdot \frac{\eta}{\delta} = \frac{2}{\ln 2} \cdot \frac{\eta}{\delta} \cdot \tau_{\rm cell-generation} \approx \frac{1}{3}\tau_{\rm cell-generation} \approx 10\ \text{min}\,. \tag{A.24}$$

For pulses whose duration is shorter than D_c, the optimal expression is zero, independent of the lactose level (i.e. the pulse amplitude L).

In a similar way, one can calculate the curve $\beta_{\rm opt}(L, D) = \beta_{\rm WT}$ and its asymptotes as shown in figure 4(*b*) in the main text.

A.6. Calculation of extrinsic noise in the lac system

Consider the case of the *lac* system activated with IPTG (schematically: $x \to R \to \beta$). The regulation consists of two stages: (i) IPTG binds to LacI, (ii) LacI unbinds from the DNA and repression is relieved. These two stages are described by the following equations:

$$R = R_T \frac{1}{1 + (x/K_{\rm IPTG-R})^2} \equiv R_T g(x) \tag{A.25}$$

$$\beta = \beta_{\max} \frac{1}{1 + (R/K_{\rm R-dna})} \equiv \beta(R). \tag{A.26}$$

The composite regulation function is given by

$$\beta(x) = \beta_{\max} \cdot \frac{1}{1 + \frac{R_T}{K_{\rm R-dna}} \frac{1}{1+(x/K_{\rm IPTG-R})^2}} \tag{A.27}$$

In order to calculate the noise we write

$$\beta = \beta(R) = \beta(R_T g(x)) = \beta_{\max} \frac{1}{1 + \frac{R_T g(x)}{K_{\rm R-dna}}}. \tag{A.28}$$

Differentiating with respect to LacI concentration (R_T) we get

$$\begin{aligned}\Delta\beta &= \frac{\mathrm{d}\beta}{\mathrm{d}R_T}\Delta R_T = \frac{\mathrm{d}\beta}{\mathrm{d}R}\frac{\mathrm{d}R}{\mathrm{d}R_T}\Delta R_T \\ &= -\beta_{\max}\frac{1}{\left(1+\frac{R_T g(x)}{K_{\mathrm{R-dna}}}\right)^2}\frac{1}{K_{\mathrm{R-dna}}}g(x)\Delta R_T \\ &= -\beta_{\max}\frac{1}{1+\frac{R_T g(x)}{K_{\mathrm{R-dna}}}}\cdot\frac{(R_T g(x)/K_{\mathrm{R-dna}})}{1+\frac{R_T g(x)}{K_{\mathrm{R-dna}}}}\cdot\frac{\Delta R_T}{R_T} \\ &= -\beta\left(1-\frac{\beta}{\beta_{\max}}\right)\cdot\frac{\Delta R_T}{R_T}. \end{aligned} \tag{A.29}$$

Thus

$$\frac{\Delta\beta}{\beta_{\max}} = \frac{\beta}{\beta_{\max}}\left(1-\frac{\beta}{\beta_{\max}}\right)\cdot\frac{\Delta R_T}{R_T}. \tag{A.30}$$

It can be seen that the noise is maximal when the system is not fully activated or fully repressed. Similar behavior can be shown numerically for induction with lactose.

A.7. The magnitude of extrinsic noise is proportional to the slope of the regulation function and the CV of the repressor concentration

Given a general regulation system activated by signal L and mediated by a transcription factor R (schematically: $L \to R \to \beta$), the regulation function $\beta(L)$ is a composite of two functions representing the two stages of regulation.

(i) The signal binds the transcription factor: $R = R_T g(L)$.
(ii) The transcription factor binds the DNA: $\beta = \beta(R)$.

Calculating the noise gives (to first order)

$$\Delta\beta \approx \frac{\mathrm{d}\beta}{\mathrm{d}R_T}\Delta R_T = \frac{\mathrm{d}\beta}{\mathrm{d}R}\frac{\mathrm{d}R}{\mathrm{d}R_T}\Delta R_T = \frac{\mathrm{d}\beta}{\mathrm{d}R}g(L)\Delta R_T. \tag{A.31}$$

Calculating the derivative of $\beta(L)$ gives

$$\frac{\mathrm{d}\beta}{\mathrm{d}L} = \frac{\mathrm{d}\beta}{\mathrm{d}R}\frac{\mathrm{d}R}{\mathrm{d}L} = \frac{\mathrm{d}\beta}{\mathrm{d}R}R_T g'(L). \tag{A.32}$$

From the last two relations we get

$$\Delta\beta \approx \frac{\mathrm{d}\beta}{\mathrm{d}L}\frac{g(L)}{g'(L)}\frac{\Delta R_T}{R_T} = \frac{\mathrm{d}\beta}{\mathrm{d}L}\frac{\Delta R_T}{R_T}\cdot B(L), \tag{A.33}$$

where $B(L) \equiv g(L)/g'(L)$ is a function of the substrate concentration L. If $g(L)$ is a Hill function with Hill coefficient n (as in the *lac* system), the noise is given by

$$\Delta\beta \approx \frac{\mathrm{d}\beta}{\mathrm{d}L}\frac{\Delta R_T}{R_T}\cdot\frac{L}{n}. \tag{A.34}$$

In our analysis we use the last form, without loss of generality.

A.8. Extrinsic noise causes a reduction in the population fitness, which is proportional to the variance of the noise

The fitness function $G(\beta, L)$ is concave, with a second derivative which is negative for any L. Taking into account the cell–cell variability $\Delta\beta$ in the production rate, one finds

$$\begin{aligned}G(\beta+\Delta\beta, L) &= G(\beta, L) + \frac{\partial}{\partial\beta}G(\beta, L)\cdot\Delta\beta \\ &+ \frac{1}{2}\frac{\partial^2}{\partial\beta^2}G(\beta, L)\cdot\Delta\beta^2 + \cdots. \end{aligned} \tag{A.35}$$

Averaging over the noise (that is, over a population of cells), and assuming that the noise is symmetric around the mean, i.e. $\langle\Delta\beta\rangle = 0$, we have

$$\begin{aligned}\langle G_{\mathrm{population}}(\beta, L)\rangle &= \frac{1}{N}\sum_{i=1}^{N} G(\beta+\Delta\beta_i, L) = \langle G(\beta+\Delta\beta, L)\rangle \\ &= G(\beta, L) - \left|\frac{1}{2}\frac{\partial^2}{\partial\beta^2}G(\beta, L)\right|\cdot\langle\Delta\beta^2\rangle + \cdots \end{aligned} \tag{A.36}$$

where N is the number of cells in the population and i is the cell index. Thus, the extrinsic noise causes reduction of the fitness, which is proportional to the noise variance $\langle\Delta\beta^2\rangle$. Substituting the fitness function and the expression for extrinsic noise derived earlier one finds

$$\begin{aligned}\langle G_{\mathrm{population}}(\beta, L)\rangle &= \left[-\eta\frac{\beta}{1-\beta/M} + \delta\beta\frac{L}{L+K_Y}\right] \\ &- \frac{\eta}{M}\frac{1}{(1-\beta/M)^3}\cdot\left(\frac{\mathrm{d}\beta}{\mathrm{d}L}\right)^2\frac{L^2}{n^2}\cdot\frac{\langle\Delta R_T^2\rangle}{R_T^2}. \end{aligned} \tag{A.37}$$

Assuming that the noise in LacI follows the non-Poissonian relation [26, 28, 32, 33]:

$$\frac{\langle\Delta R_T^2\rangle}{R_T^2} = \frac{b}{R_T} \tag{A.38}$$

where R_T is the number of repressor molecules per cell and b is the burst size (taken, without loss of generality[9], to be $b = 50$), we get

$$\begin{aligned}\langle G_{\mathrm{population}}(\beta, L)\rangle &= \left[-\eta\frac{\beta}{1-\beta/M} + \delta\beta\frac{L}{L+K_Y}\right] \\ &- \frac{\eta}{M}\frac{1}{(1-\beta/M)^3}\frac{L^2}{n^2}\cdot\frac{b}{R_T}\cdot\left(\frac{\mathrm{d}\beta}{\mathrm{d}L}\right)^2. \end{aligned} \tag{A.39}$$

The generalized fitness function is found by integrating over all lactose concentrations, taking into account the probability for having each concentration in the environment. Furthermore, we also add the cost of constitutively producing LacI:

$$\begin{aligned}J\{\beta(L)|P\} &= -\eta_R R_T + \int_{L=0}^{\infty}\mathrm{d}L\cdot P(L)\cdot G_{\mathrm{population}}(\beta, L) \\ &= -\eta_R R_T + \int_{L=0}^{\infty}\mathrm{d}L\cdot P(L)\cdot\left[-\eta\frac{\beta}{1-\beta/M} + \delta\beta\frac{L}{L+K_Y}\right. \\ &\left. - \frac{\eta}{M}\frac{1}{(1-\beta/M)^3}\frac{L^2}{n^2}\cdot\frac{b}{R_T}\cdot\left(\frac{\mathrm{d}\beta}{\mathrm{d}L}\right)^2\right] \\ &= -\eta_R R_T + \int_{L=0}^{\infty}\mathrm{d}L\cdot P(L)\cdot\left[-\eta\frac{\beta}{1-\beta/M} + \delta\beta\frac{L}{L+K_Y}\right. \\ &\left. - \frac{C(\beta, L)}{R_T}\cdot\left(\frac{\mathrm{d}\beta}{\mathrm{d}L}\right)^2\right] \end{aligned} \tag{A.40}$$

where

$$C(\beta, L) = \frac{\eta}{M}\frac{1}{(1-\beta/M)^3}\frac{L^2}{n^2}b\times\frac{1}{10^9}. \tag{A.41}$$

Note that in the final form of the generalized fitness function, the amount of repressor (R_T) is expressed in molars rather than the number of molecules. In order to convert the number of molecules per cell to concentrations we use: R_T (molecules/cell) $= 10^9 \times R_T$ (molar).

[9] Similar results were obtained using a wide range of burst sizes, e.g. $10 \leqslant b \leqslant 500$.

Using the generalized fitness function, it is possible to derive the functional form of the regulation function of the *lac* system from principles of evolutionary cost and benefit, without restricting ourselves to the model described earlier. We only need to assume a general scheme of a gene regulation system consisting of (i) an input signal L, (ii) a transcription factor R (activator/repressor), and (iii) an operon producing the required protein in response to the external signal according to the regulation function $F(L)$.

A.9. Solving variational problems with simple inequality constraints using gradient-based methods

Assume that $F(L)$ represents an arbitrary regulation function. When given the lactose probability distribution $P(L)$ and the repressor concentration R_T, the generalized fitness function is a *functional* $J\{F(L)\}$, mapping the regulation function $F(L)$ to its fitness. Finding the optimal regulation function with maximal fitness is a variational problem. However, there are also physical constraints on the regulation function: $F(L) \geqslant 0$ and $F(L) < M$. Such constraints are relevant to our problem of optimizing the generalized fitness function since below $L_c = K_y(\delta/\eta - 1)^{-1} \approx 60\ \mu$M the optimal production rate without noise is formally negative (see the main paper). Variational problems with inequality constraints are rather hard to solve [54, 55]. However, if the inequality constraints are simple, as in our case, then the problem can be solved numerically using a gradient-based method demonstrated below.

The gradient algorithm. In the gradient method, we sample the interval $[0, L]$ at N points: $L_1, \ldots, L_N$, where $L_n = (L/N)n$. We also represent the functional in discrete form: $J(F_1, F_2, \ldots, F_N)$, where $F_n \equiv F(L_n)$ and $\mathrm{d}L \equiv L/N$. Thus, integral is replaced by a discrete sum, and the functional $J\{F(L)\}$ becomes an N-dimensional function $J(F_1, F_2, \ldots, F_N)$ with N variables ($J : \mathbb{R}^N \to \mathbb{R}$).

In order to find the maximal point $\mathbf{F}_{\max} = \left(F_1^{(\max)}, F_2^{(\max)}, \ldots, F_N^{(\max)}\right)$, we start with an initial point $\mathbf{F}_0 = \left(F_1^{(0)}, F_2^{(0)}, \ldots, F_N^{(0)}\right)$. We then calculate the gradient $\nabla J = \left(\frac{\partial J}{\partial F_1}, \frac{\partial J}{\partial F_2}, \ldots, \frac{\partial J}{\partial F_N}\right)$ and add a small vector in the direction of the gradient. We repeat the process, i.e., at each iteration we add $\mathbf{F}_{\mathrm{i}+1} = \mathbf{F}_{\mathrm{i}} + \varepsilon \frac{\nabla J}{||\nabla J||}$, where ε is a small number chosen in advance. After a predetermined number of steps the algorithm ends. If the number of steps is large enough, the approximate maximum $\mathbf{F}_{\max}$ of $J(\mathbf{F})$ will be reached, and ∇J will be close to zero. Note that the equation $\nabla J = 0$ corresponds to the Euler–Lagrange equations of variational calculus.

In the case of inequality constraints, for example: $F(L) \geqslant 0$, we follow the same iterative procedure. However, if the constraint is violated after some iteration, i.e., if $F_n < 0$ for some $1 \leqslant n \leqslant N$, then we set $F_n = 0$. Thus, we externally prevent the function $F(L)$ from violating the constraints. A sketch of the gradient-based algorithm is shown in figure A4.

We note that there exist faster converging optimization algorithms, such as the Newton method, which involves dividing the gradient by the negative inverse of the fitness curvature (the Hessian matrix): $\mathbf{F}_{\mathrm{i}+1} = \mathbf{F}_{\mathrm{i}} - J'(\mathbf{F}_{\mathrm{i}})/J''(\mathbf{F}_{\mathrm{i}})$. However, these methods involve time-consuming computations.

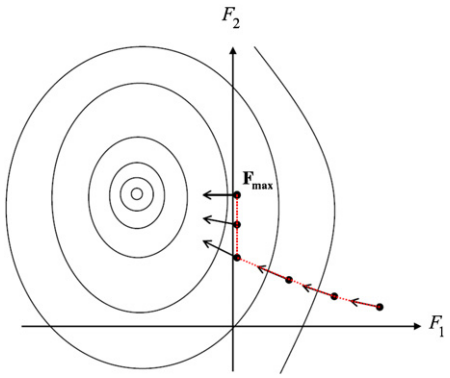

Figure A4. An illustration of the gradient based algorithm for solving optimization problems with the inequality constraints $F_1, F_2, \ldots, F_N \geqslant 0$. In each iteration the solution $\mathbf{F} = (F_1, F_2, \ldots, F_N)$ (dots) advances in the direction of the gradient $\nabla J = (\frac{\partial J}{\partial F_1}, \frac{\partial J}{\partial F_2}, \ldots, \frac{\partial J}{\partial F_N})$ (arrows), thus advancing toward the maximum of the fitness function $J(F_1, F_2, \ldots, F_N)$. However, if an inequality constraint is violated, e.g. $F_1 < 0$, the solution is fixed by setting $F_1 = 0$. Note that the gradient does not vanish at the maximal point $\mathbf{F}_{\max}$.

The examples presented below are demonstrated in movies available at stacks.iop.org/PhysBio/4/229 that can also be found at http://www.weizmann.ac.il/mcb/UriAlon/people/Tomer/TomerMovies.html

Example 1. Finding the shortest path between two points.

In order to find the trajectory $y(x)$ that minimizes the path length between two points in the X–Y plane, we can write the functional:

$$J\{y(x)\} = \int_{x_1}^{x_1} [1 + (y')^2]\,\mathrm{d}x \equiv \int_{x_1}^{x_1} L(x, y, y')\,\mathrm{d}x. \qquad \text{(A.42)}$$

The Euler–Lagrange equations are $\frac{\partial L}{\partial y} - \frac{\mathrm{d}}{\mathrm{d}x}\left(\frac{\partial L}{\partial y'}\right) = -2y'' = 0$, which gives the straight line solution $y(x) = ax + b$ where a and b are constants.

In discrete form the functional is given by

$$J(y_1, \ldots, y_N) = \sum_{n=1}^{N-1} \left[1 + \left(\frac{y_{n+1} - y_n}{\mathrm{d}x}\right)^2\right] \cdot \mathrm{d}x, \qquad \text{(A.43)}$$

where $y_n \equiv y(x_n)$.

In order to find the minimal point $\mathbf{y_{min}} = \left(y_1^{(\min)}, y_2^{(\min)}, \ldots, y_N^{(\min)}\right)$ we calculate the gradient $\nabla J = \left(\frac{\partial J}{\partial y_1}, \frac{\partial J}{\partial y_2}, \ldots, \frac{\partial J}{\partial y_N}\right)$. We assume that the endpoints are fixed, so we have $\frac{\partial J}{\partial y_1} = \frac{\partial J}{\partial y_N} = 0$. For $1 < n < N$ we have

$$\frac{\partial J}{\partial y_n} = 2\left(\frac{y_{n+1} - y_n}{\mathrm{d}x}\right)\left(-\frac{1}{\mathrm{d}x}\right) + 2\left(\frac{y_n - y_{n-1}}{\mathrm{d}x}\right)\left(\frac{1}{\mathrm{d}x}\right) = -2\frac{[(y_{n+1} - y_n)/\mathrm{d}x - (y_n - y_{n-1})/\mathrm{d}x]}{\mathrm{d}x}, \qquad \text{(A.44)}$$

which is a discrete form of the Euler–Lagrange expression.

Starting from some initial solution $\mathbf{y_0} = \left(y_1^{(0)}, y_2^{(0)}, \ldots, y_N^{(0)}\right)$ (with the endpoints y_1 and y_N set according

to the boundary conditions), the solution can be found numerically by the gradient algorithm (see movie no. 1 available at stacks.iop.org/PhysBio/4/229).

Example 2. Finding the shortest path between two points with an impassable barrier in between.

Assume an impassable barrier in the form of a circle of radius R around the origin. All values of $y(x)$ must obey $x^2 + y^2 \geqslant R^2$, or $y^2 \geqslant R^2 - x^2$. This imposes an inequality constraint on the variational problem. However, the constrained variational problem can be solved numerically by the gradient method as before. At each iteration, if the constraint is violated, i.e. if for some y_n ($n \in \{1, \ldots, N\}$) we have $y_n < \sqrt{R^2 - x_n^2}$, we externally force the constraint by setting $y_n = \sqrt{R^2 - x_n^2}$.

The numerical solution for this case can be seen in movie no. 2 available at stacks.iop.org/PhysBio/4/229. Solutions conform to analytical results [55]. Note that gradient does not vanish at the maximal point in places where the constraint is forced ($-R < x < R$), as demonstrated in figure A4.

Example 3. Finding the optimal regulation function for a given lactose distribution $P(L)$.

A discrete form of the generalized fitness function is given by

$$J(F_1, F_2, \ldots, F_N) = -\eta_R R_T + \sum_{n=1}^{N-1} \mathrm{d}L \cdot P(L_n) \cdot \left[-\eta \frac{F_n}{1 - F_n/M} + \delta F_n \frac{L_n}{L_n + K_Y} - \frac{C(F_n, L_n)}{R_T} \cdot \left(\frac{F_{n+1} - F_n}{\mathrm{d}L} \right)^2 \right] + \mathrm{d}L \cdot P(L_N) \cdot \left[-\eta \frac{F_N}{1 - F_N/M} + \delta F_n \frac{L_N}{L_N + K_Y} \right]. \quad \text{(A.45)}$$

The gradient $\nabla J = \left(\frac{\partial J}{\partial F_1}, \frac{\partial J}{\partial F_2}, \ldots, \frac{\partial J}{\partial F_N} \right)$ is calculated by taking partial derivatives with respect to $F_1, F_2, \ldots, F_N$. Note that the first and last components of the gradient (i.e., the endpoints $n = 1$ and $n = N$) are different than all other components.

We solve using the gradient method as before, while forcing the constraints: $F_n \geqslant 0$ and $F_n < M$ for any $n \in \{1, \ldots, N\}$. At each iteration, if $F_n < 0$ we set $F_n = 0$, and if $F_n > M$ we set $F_n = 0.99\,M$. Results can be seen in movie no. 3 available at stacks.iop.org/PhysBio/4/229 and in figure 7(*a*) in the main text.

Glossary

The lac operon: a set of genes found in the bacterium *E. coli* (and other enteric bacteria) which are activated in the presence of the sugar lactose in the environment. The operon consists of three genes, *lacZ*, *lacY* and *lacA*, which are responsible for lactose catabolism and transport. The *lac* operon is regulated by a constitutively expressed repressor protein (LacI) which binds tightly to the DNA in its promoter region and thus prevents transcription of the lactose utilization proteins. When lactose is introduced into the growth medium, the *lac* repressor is modified such that it cannot bind to the DNA, thus allowing production of LacZ, LacY and LacA at high levels.

Cost–benefit evolutionary analysis: a theory for predicting the evolutionary outcome given the environmental conditions. In microbiology, it is assumed that the design of a bacterial species (e.g. gene expression level, shape of regulation function) is selected by evolution such as to maximize the fitness (i.e. growth rate) of the bacterial population of this species. The selected design is an optimal balance between cost and benefit. The cost quantifies the growth burden on cell resources due to production of a protein or due to maintaining a certain regulatory mechanism. The benefit is the growth advantage conferred by the function of this mechanism. For example, maintaining high transcription factor levels constitutively reduces noise in gene expression (benefit) while straining the limited resources of the cell (cost).

Noise in gene expression: cell–cell variability of protein or mRNA levels in genetically identical cells due to the stochastic mechanism of gene expression. Strong noise in a specific gene may also originate from random variations in an upstream regulator, especially if the regulator is expressed in low levels.

Feed-forward loop (FFL): a regulatory gene circuit in which two regulators X and Y both regulate a third gene Z, and also X regulates Y. Previous work has shown that FFL's serve as pulse generators, response accelerators, or sign-sensitive delay elements that can protect against an unwanted response to fluctuating inputs.

References

[1] Elena S F and Lenski R E 2003 Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation *Nat. Rev. Genet.* **4** 457–69

[2] Orr H A 2005 The genetic theory of adaptation: a brief history *Nat. Rev. Genet.* **6** 119–27

[3] Ibarra R U, Edwards J S and Palsson B O 2002 *Escherichia coli* K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth *Nature* **420** 186–9

[4] Koch A L 1983 The protein burden of lac operon products *J. Mol. Evol.* **19** 455–62

[5] Dekel E and Alon U 2005 Optimality and evolutionary tuning of the expression level of a protein *Nature* **436** 588–92

[6] Dekel E, Mangan S and Alon U 2005 Environmental selection of the feed-forward loop circuit in gene-regulation networks *Phys. Biol.* **2** 81–8

[7] Thattai M and van Oudenaarden A 2004 Stochastic gene expression in fluctuating environments *Genetics* **167** 523–30

[8] Wolf D M, Vazirani V V and Arkin A P 2005 Diversity in times of adversity: probabilistic strategies in microbial survival games *J. Theor. Biol.* **234** 227–53

[9] Jacob F and Monod J 1961 Genetic regulatory mechanisms in the synthesis of proteins *J. Mol. Biol.* **3** 318–56

[10] Muller-Hill B 1996 *The Lac Operon: A Short History of a Genetic Paradigm* (Berlin: Walter de Gruyter) p 207

[11] Alon U 2006 *An Introduction to Systems Biology: Design Principles of Biological Circuits* (Mathematical and Computational Biology Series) 1st edn (Boca Raton, FL: CRC Press)

[12] Mayo A E, Setty Y, Shavit S, Zaslaver A and Alon U 2006 Plasticity of the cis-regulatory input function of a gene *PLoS Biol.* **4** e45

[13] Setty Y, Mayo A E, Surette M G and Alon U 2003 Detailed map of a cis-regulatory input function *Proc. Natl Acad. Sci. USA* **100** 7702–7

[14] Wang X G, Olsen L R and Roderick S L 2002 Structure of the *lac operon* galactoside acetyltransferase *Structure* **10** 581–8

[15] Andrews K J and Lin E C 1976 Thiogalactoside transacetylase of the lactose operon as an enzyme for detoxification *J. Bacteriol.* **128** 510–3

[16] Santillan M and Mackey M C 2004 Influence of catabolite repression and inducer exclusion on the bistable behavior of the *lac operon Biophys. J.* **86** 1282–92

[17] Yildirim N and Mackey M C 2003 Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data *Biophys. J.* **84** 2841–51

[18] Wong P, Gladney S and Keasling J D 1997 Mathematical model of the *lac operon*: inducer exclusion, catabolite repression, and diauxic growth on glucose and lactose *Biotechnol. Prog.* **13** 132–43

[19] Yagil G and Yagil E 1971 On the relation between effector concentration and the rate of induced enzyme synthesis *Biophys. J.* **11** 11–27

[20] Vilar J M and Leibler S 2003 DNA looping and physical constraints on transcription regulation *J. Mol. Biol.* **331** 981–9

[21] Mangan S and Alon U 2003 Structure and function of the feed-forward loop network motif *Proc. Natl Acad. Sci. USA* **100** 11980–5

[22] Savageau M A 2001 Design principles for elementary gene circuits: elements, methods, and examples *Chaos* **11** 142–59

[23] Kussell E and Leibler S 2005 Phenotypic diversity, population growth, and information in fluctuating environments *Science* **309** 2075–8

[24] Kaern M, Elston T C, Blake W J and Collins J J 2005 Stochasticity in gene expression: from theories to phenotypes *Nat. Rev. Genet.* **6** 451–64

[25] Sigal A, Milo R, Cohen A, Geva-Zatorsky N, Klein Y, Liron Y, Rosenfeld N, Danon T, Perzov N and Alon U 2006 Variability and memory of protein levels in human cells *Nature* **444** 643–6

[26] Bar-Even A, Paulsson J, Maheshri N, Carmi M, O'Shea E, Pilpel Y and Barkai N 2006 Noise in protein expression scales with natural protein abundance *Nat. Genet.* **38** 636–43

[27] Raser J M and O'Shea E K 2005 Noise in gene expression: origins, consequences, and control *Science* **309** 2010–3

[28] Thattai M and van Oudenaarden A 2001 Intrinsic noise in gene regulatory networks *Proc. Natl Acad. Sci. USA* **98** 8614–9

[29] Volfson D, Marciniak J, Blake W J, Ostroff N, Tsimring L S and Hasty J 2006 Origins of extrinsic variability in eukaryotic gene expression *Nature* **439** 861–4

[30] Cai L, Friedman N and Xie X S 2006 Stochastic protein expression in individual cells at the single molecule level *Nature* **440** 358–62

[31] Elowitz M B, Levine A J, Siggia E D and Swain P S 2002 Stochastic gene expression in a single cell *Science* **297** 1183–6

[32] Friedman N, Cai L and Xie X S 2006 Linking stochastic dynamics to population distribution: an analytical framework of gene expression *Phys. Rev. Lett.* **97** 168302

[33] McAdams H H and Arkin A 1997 Stochastic mechanisms in gene expression *Proc. Natl Acad. Sci. USA* **94** 814–9

[34] Swain P S, Elowitz M B and Siggia E D 2002 Intrinsic and extrinsic contributions to stochasticity in gene expression *Proc. Natl Acad. Sci. USA* **99** 12795–800

[35] Rosenfeld N, Elowitz M B and Alon U 2002 Negative autoregulation speeds the response times of transcription networks *J. Mol. Biol.* **323** 785–93

[36] Rosenfeld N, Young J W, Alon U, Swain P S and Elowitz M B 2005 Gene regulation at the single-cell level *Science* **307** 1962–5

[37] Dong H, Nilsson L and Kurland C G 1995 Gratuitous overexpression of genes in *Escherichia coli* leads to growth inhibition and ribosome destruction *J. Bacteriol.* **177** 1497–504

[38] Shen-Orr S S, Milo R, Mangan S and Alon U 2002 Network motifs in the transcriptional regulation network of *Escherichia coli Nat. Genet.* **31** 64–8

[39] Kalir S, Mangan S and Alon U 2005 A coherent feed-forward loop with a SUM input function prolongs flagella expression in *Escherichia coli Mol. Syst. Biol.* **1** E1–6

[40] Mangan S, Itzkovitz S, Zaslaver A and Alon U 2006 The incoherent feed-forward loop accelerates the response-time of the gal system of *Escherichia coli J. Mol. Biol.* **356** 1073–81

[41] Mangan S, Zaslaver A and Alon U 2003 The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks *J. Mol. Biol.* **334** 197–204

[42] Alon U 2007 Network motifs: theory and experimental approaches *Nat. Rev. Genet.* **8** 450–61

[43] Debarbouille M and Schwartz M 1979 The use of gene fusions to study the expression of malT the positive regulator gene of the maltose regulon *J. Mol. Biol.* **132** 521–34

[44] Richet E and Raibaud O 1987 Purification and properties of the MalT protein, the transcription activator of the *Escherichia coli* maltose regulon *J. Biol. Chem.* **262** 12647–53

[45] Casadaban M J 1976 Regulation of the regulatory gene for the arabinose pathway, araC *J. Mol. Biol.* **104** 557–66

[46] Hahn S and Schleif R 1983 In vivo regulation of the *Escherichia coli* araC promoter *J. Bacteriol.* **155** 593–600

[47] Kolodrubetz D and Schleif R 1981 Identification of araC protein and two-dimensional gels, its *in vivo* instability and normal level *J. Mol. Biol.* **149** 133–9

[48] Geanacopoulos M and Adhya S 1997 Functional characterization of roles of GalR and GalS as regulators of the gal regulon *J. Bacteriol.* **179** 228–34

[49] Kar S and Adhya S 2001 Recruitment of HU by piggyback: a special role of GalR in repressosome assembly *Genes Dev.* **15** 2273–81

[50] Shinar G, Dekel E, Tlusty T and Alon U 2006 Rules for biological regulation based on error minimization *Proc. Natl Acad. Sci. USA* **103** 3999–4004

[51] Zaslaver A, Bren A, Ronen M, Itzkovitz S, Kikoin I, Shavit S, Liebermeister W, Surette M G and Alon U 2006 A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli Nat. Method* **3** 623–8

[52] Balagadde F K, You L, Hansen C L, Arnold F H and Quake S R 2005 Long-term monitoring of bacteria undergoing programmed population control in a microchemostat *Science* **309** 137–40

[53] Ozbudak E M, Thattai M, Lim H N, Shraiman B I and Van Oudenaarden A 2004 Multistability in the lactose utilization network of *Escherichia coli Nature* **427** 737–40

[54] Valentine F A 1946 *The problem of Lagrange with differential inequalities as added side conditions Contributions to the Calculus of Variations (1933–1937)* ed G A Bliss (Chicago: University of Chicago Press) pp 407–48

[55] Desai J P 1998 Motion planning and control of cooperative robotic systems *Mechanical Engineering and Applied Mechanics* (Philadelphia, PA: University of Pennsylvania) p 180