INSTITUTE OF PHYSICS PUBLISHING PHYSICAL BIOLOGY
Phys. Biol. **2** (2005) 81–88 doi:10.1088/1478-3975/2/2/001

Environmental selection of the feed-forward loop circuit in gene-regulation networks

Erez Dekel, Shmoolik Mangan and Uri Alon

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

E-mail: urialon@weizmann.ac.il

Received 23 December 2004
Accepted for publication 31 March 2005
Published 28 April 2005
Online at stacks.iop.org/PhysBio/2/81

Abstract

Gene-regulation networks contain recurring elementary circuits termed network motifs. It is of interest to understand under which environmental conditions each motif might be selected. To address this, we study one of the most significant network motifs, a three-gene circuit called the coherent feed-forward loop (FFL). The FFL has been demonstrated theoretically and experimentally to perform a basic information-processing function: it shows a delay following ON steps of an input inducer, but not after OFF steps. Here, we ask under what environmental conditions might the FFL be selected over simpler gene circuits, based on this function. We employ a theoretical cost–benefit analysis for the selection of gene circuits in a given environment. We find conditions that the environment must satisfy in order for the FFL to be selected over simpler circuits: the FFL is selected in environments where the distribution of the input pulse duration is sufficiently broad and contains both long and short pulses. Optimal values of the biochemical parameters of the FFL circuit are determined as a function of the environment such that the delay in the FFL blocks deleterious short pulses of induction. This approach can be generally used to study the evolutionary selection of other network motifs.

Introduction

Biological networks contain network motifs: connectivity patterns that recur in many different systems [1–3]. Network motifs may be readily detected because they appear much more often than in randomized networks [1–3]. Transcription regulation networks show several highly significant network motifs. Each of the network motifs in transcription networks has been demonstrated to carry out a basic information-processing function [4].

One of the most significant network motifs is the feed-forward loop (FFL), in which a transcription factor X regulates a second transcription factor Y, and both jointly regulate gene Z (or several genes $Z_1, \ldots, Z_n$) (figure 1(a)) [1]. The FFL appears in diverse organisms including *E. coli* [1–3, 5, 6], *B. subtilis* [3, 7], yeast [2, 5, 8, 9], *C. elegans* [6], fruit-fly [3], sea urchin [3,10] and humans [11]. For example, sporulation of *B. subtilis* is controlled by a transcriptional network made of several feed-forward loops [7]. Evolution appears to have independently converged on this motif in different organisms as well as in different systems within the same organism [6, 12].

The dynamical behavior of the FFL depends on the nature of the regulatory interactions (activation or repression) between X, Y and Z, and on the cis-regulatory input function, that integrates the effects of X and Y on Z [13–15]. A common input function is an AND-gate in which both X and Y are needed to activate Z [5, 6]. The functions of the various possible FFL variants have been analyzed [5, 6].

The most common FFL configuration, called the coherent type-1 FFL [5], has three activation regulations (figure 1(a)). This circuit functions as a sign-sensitive delay element [1, 5, 6]: following a step-like addition of the stimulus of X, S_x, the output gene Z is activated at a delay. The delay is due to the fact that Y must accumulate and cross its activation threshold in order to activate Z. No delay occurs, however, upon a step-like removal of the stimulus S_x. This is because only one input of the AND-gate needs to go off for Z to be deactivated.

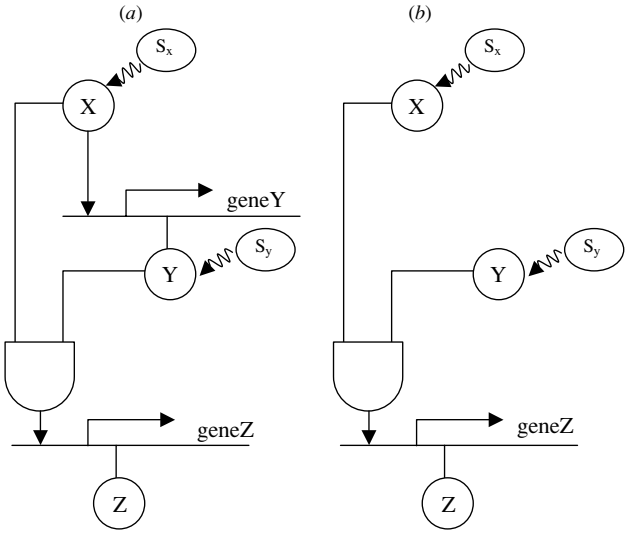

Figure 1. Feed-forward loop (FFL) and simple-AND regulation circuits. (*a*) Feed-forward loop, where X activates Y and both jointly activate gene Z in an AND-gate fashion. The inducers are S_x and S_y. In the *ara* system, for example, X = CRP, Y = araC, Z = araBAD, S_x = cAMP and S_y = L-arabinose. (*b*) A simple-AND-gate regulation-circuit, where X and Y activate gene Z. In the *lac* system, for example, Y = lacI is a repressor that is induced by S_y = lactose, X = CRP and S_x = cAMP.

This function can be viewed as a persistence detector: Z is expressed only in response to sufficiently long pulses of the input, S_x, whereas rapid deactivation of Z expression occurs when S_x is removed. These dynamical features have been experimentally demonstrated in the FFL that regulates the L-arabinose utilization system of *E. coli* [6].

Not all systems regulated by two inputs exhibit the FFL: for example, the lactose system of *E. coli* [14, 16, 17] is a simple-AND-gate structure, where X (CRP) does not regulate Y (LacI) (figure 1(*b*)). The FFL is found in other *E. coli* sugar systems with the same X (CRP), such as the arabinose, fucose and maltose systems [18–23]. About 40% of the *E. coli* operons known to be regulated by two inputs participate in a FFL [1].

What determines why the FFL is selected in some systems and not others? It is known that the arrows in regulatory networks can rapidly change over evolutionary timescales [9, 12]. For example, it only takes a few point mutations in the binding site of X in the promoter of Y to abolish the interaction $X \rightarrow Y$. Of the three arrows in the FFL, two are essential for maintaining the circuits' AND-gate decision-making logic. These are the arrows $X \rightarrow Z$ and $Y \rightarrow Z$. The third arrow, $X \rightarrow Y$, can be removed without disrupting the AND-gate logic of the circuit. Therefore, we can ask, what preserves the regulation of Y by X in the FFL against mutations that would rapidly abolish this interaction?

To address this, we use a theoretical evolutionary approach to test the hypothesis that the dynamical properties of the FFL convey an advantage to the cell under certain environmental conditions. Evolutionary analysis based on optimality principles is a classic approach [24]. Examples have been presented for several design features in biological regulatory and metabolic systems [12, 25–48]. Pioneering studies include rules for determining the mode of regulation based on demand theory [25, 31, 37]; the structure of the pentose–phosphate pathway as an evolutionary game minimizing the number of reaction steps [28, 30]; rules for optimal design of metabolic pathways for maximal efficiency and rapid responses while minimizing total enzyme production [26, 27, 29, 30, 32, 35, 38, 40, 42, 47]; mathematically controlled comparison of different designs for genetic switches [12, 31, 34–37, 45]; analysis of optimal genome arrangement in phage [49]; and global optimization of metabolic fluxes [30, 38, 42, 44, 50].

Here, we present a simple model for the selection of the FFL, based on a cost–benefit analysis of protein action in a changing environment. We find analytical conditions for FFL selection in terms of the environmental input distribution. This may provide an explanation why FFL is found in some systems and not in others. It also provides insight into the selected values of the biochemical parameters of the FFL in a given environment.

Results

Cost–benefit analysis of a simple gene-regulation circuit

We analyze a gene-regulation system with two inputs that control expression of gene Z. We begin with regulation by a simple-AND circuit (figure 1(*b*)) and consider the FFL in the next section. Production of protein Z is ON at a constant rate β in the presence of both input inducers S_x and S_y, and otherwise zero.

We consider the effects of production of protein Z on the growth rate of the cells. The cost of Z production entails a reduction in growth rate $-\eta\beta$, where β is the rate of production of Z and η is the reduction in growth rate per Z molecule produced[1].

On the other hand, the action of the Z gene-product conveys an advantage to the cells. This advantage is described by $\delta f(Z)$, the increase in growth rate due to the action of Z. $f(Z)$ is typically an increasing function of Z that saturates at high values of Z.

An example is the arabinose sugar catabolism system of *E. coli*. Here, $\delta f(Z)$ represents the increase in growth rate due to the energy and carbon supplied to the cells by catabolism of the sugar S_y = arabinose. The input signal S_x in the arabinose system is cAMP, a signaling molecule produced in the cell upon glucose starvation. In the arabinose system, both S_x = cAMP and S_y = arabinose need to be present for benefit, because of catabolite-exclusion in the absence of S_x, e.g. in the presence of glucose. In this system, $f(Z)$ is the rate at which Z breaks down sugar S_y. This rate can be described by Michaelis–Menten enzyme kinetics: $\delta f(Z) = \delta_0 v S_y Z/(K+Z)$, where K

[1] Typically, the costs for the production of the transcription factors X and Y are negligible compared to the production cost of the effector protein Z [50], since transcription factors are typically produced in far fewer copies per cell than enzymes or structural proteins. If Y costs are not negligible, the advantage of FFL over simple-AND increases, because the FFL prevents unneeded Y production. Y production costs are included in the detailed model in the appendix.

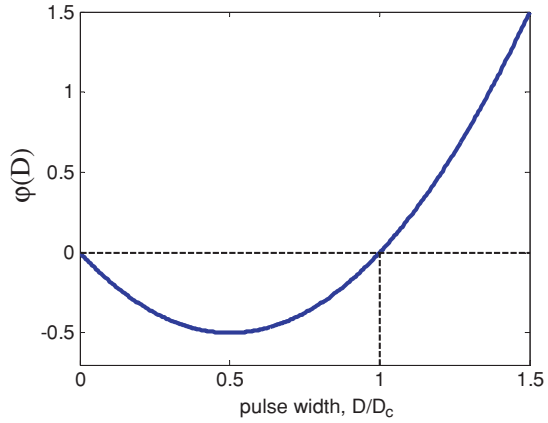

Figure 2. Fitness (integrated growth rate) of simple regulation during short pulses of inputs S_x and S_y. Fitness is negative for $D < D_c$.

is the Michaelis constant of the enzyme, v is the rate at which S_y is metabolized and δ_0 is the increase in growth rate per sugar molecule metabolized[2].

The overall effect of Z on the growth maximal rate is the sum of the cost and benefit [25, 30]:

$$g = -\eta\beta + \delta f(Z). \tag{1}$$

We now consider a pulse of activation, in which both S_x and S_y are present at saturating levels for a pulse of duration D. The growth of cells with a simple-AND circuit, integrated over time D, is given by

$$\varphi(D) = \int_0^D g(t)\,\mathrm{d}t = -\eta\beta D + \int_0^D \delta f(Z)\,\mathrm{d}t. \tag{2}$$

When the pulse begins, protein Z begins to be produced at rate β, and degraded or diluted out by cell growth at rate α [51]. The dynamics of Z concentration are given by

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \beta - \alpha Z \tag{3}$$

resulting in an exponential convergence to steady-state $Z_\mathrm{m} = \beta/\alpha$

$$Z(t) = Z_\mathrm{m}(1 - \mathrm{e}^{-\alpha t}). \tag{4}$$

This solution is in good agreement with high-resolution gene expression measurements [51].

For long pulses ($D\alpha \gg 1$), Z is saturated $Z = Z_\mathrm{m}$, and has a net positive effect on cell growth

$$\varphi(D) = -\eta\beta D + \delta f(Z_\mathrm{m})D \tag{5}$$

provided that the benefit of Z exceeds its production costs $\delta f(Z_\mathrm{m}) > \beta\eta$.

Short pulses, however, can have a deleterious effect on growth. To see this, consider short pulses such that $D\alpha \ll 1$. In this case $Z(t) \sim \beta t$ and using the series expansion $f(Z) \sim f'Z$, the integrated growth rate is (figure 2)

$$\varphi(D) = \int_0^D (-\eta\beta + \delta f'\beta t)\,\mathrm{d}t = -\eta\beta D + \delta f'\beta\frac{D^2}{2}. \tag{6}$$

[2] The Michaelis–Menten term applies to the case where S_y is saturating. More generally, the quadratic form of $f(Z)$, which includes sub-saturating S_y, is described in the appendix.

Growth is reduced ($\varphi(D) < 0$) for pulses shorter than a critical pulse duration D_c (figure 2)

$$D_c = \frac{2\eta}{\delta f'}. \tag{7}$$

Hence, short pulses are deleterious. Simple regulation leads to reduction in growth in environments with short pulses, even though Z confers a net advantage for sufficiently long input pulses (figure 3(*a*)).

Cost–benefit analysis of the FFL gene circuit

In the FFL (figure 1(*a*)), upon a pulse of S_x, the transcription factor Y begins to be produced

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \beta_y - \alpha_y Y \tag{8}$$

and exponentially converges to its steady-state level $Y_m = \beta_y/\alpha_y$

$$Y(t) = Y_\mathrm{m}(1 - \mathrm{e}^{-\alpha_y t}). \tag{9}$$

Gene Z in the FFL is regulated in an AND-gate fashion by X and Y. Therefore, to activate Z, Y needs to accumulate to levels sufficient to bind the Z promoter and cause activation of transcription. A simple description of regulation of Z by Y, allowing analytical solution of the dynamics, is threshold regulation, where Z is produced at rate β when $Y > T_Y$, and not produced when $Y < T_Y$. Many genes are indeed regulated with sharp regulation functions that resemble threshold regulation [14, 15, 52]. Other, less sharp, regulation functions yield the same qualitative results, as discussed in the appendix.

Thus, gene Z is only activated at a delay, at time $t = \tau$ when Y reaches its activation threshold, $Y(\tau) = T_Y$. The delay, τ, can be found from equation (9):

$$\tau = \alpha_y^{-1}\ln\left(\frac{1}{1 - T_Y/Y_\mathrm{m}}\right). \tag{10}$$

This equation relates the magnitude of the delay in Z expression to the biochemical parameters of protein Y. Typical parameter values in bacteria yield delays of the order of 1–100 min. The delay in the FFL can in principle be tuned to optimal values by mutations that change these biochemical parameters. The delay acts to filter out pulses that are shorter than τ (figure 3(*b*)). This avoids the reduction in growth for short pulses:

$$\varphi(D) = 0 \qquad \text{for} \quad D < \tau. \tag{11}$$

However, the filtering of short pulses has a disadvantage, because during long pulses, Z is produced at a delay and misses some of the potential benefit of the pulse (figure 3(*b*)). To assess whether the FFL confers a net advantage to the cells, relative to simple regulation, requires analysis of the distribution of pulses in the environment.

Conditions for FFL selection

The environment of the cell can be characterized by the probability distribution of the duration of input pulses, $P(D)$.

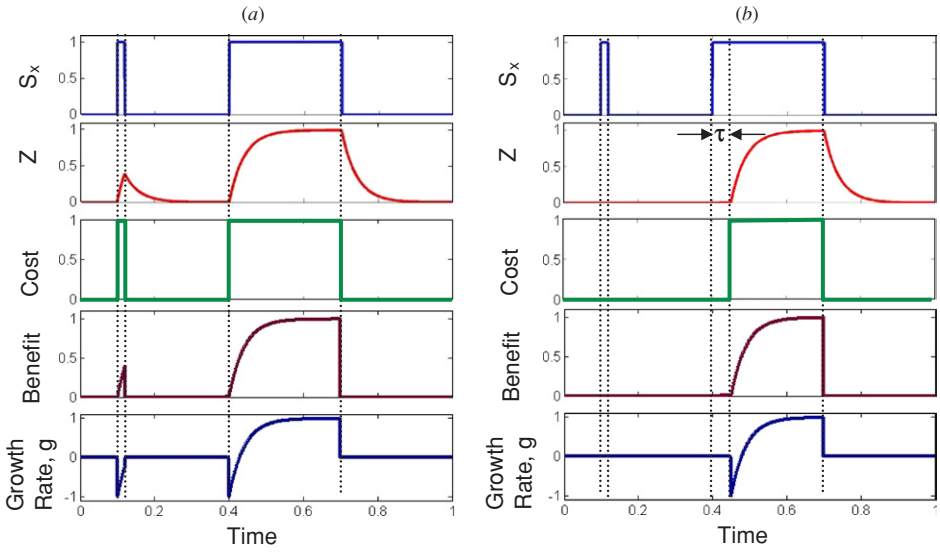

Figure 3. Dynamics of gene expression and growth rate in a short, non-beneficial pulse and a long pulse of S_x and S_y. (*a*) Simple regulation shows a growth deficit for both pulses, (*b*) FFL filters out the short pulse, but has reduced benefit during the long pulse. The figure shows (top to bottom): (1) Pulse of S_x and S_y. (2) Dynamics of Z expression. Z is turned on after a delay τ ($\tau = 0$ in the case of simple regulation), and approaches its steady-state level Z_m. (3) Normalized production cost (reduction in growth rate) due to the production load of Z. Cost begins after the delay τ. (4) Normalized growth rate advantage (benefit) from the action of gene product Z. (5) Net normalized growth rate.

We assume for simplicity that the pulses are far apart, so that the system starts each pulse from zero initial Z levels (and Y levels in the case of the FFL). In this case, the overall fitness can be found by integrating the fitness $\varphi(D)$ over the pulse distribution. For simple-AND circuits,

$$\Phi_1 = \int_0^\infty P(D)\varphi(D)\,\mathrm{d}D. \tag{12}$$

For FFL circuits, production starts after a delay τ. Pulses shorter than τ result in no Z production and $\varphi(D < \tau) = 0$. Long pulses begin to be utilized after a delay τ, so that their duration is effectively $D - \tau$ (figure 3(*b*)), resulting in

$$\Phi_2 = \int_\tau^\infty P(D)\varphi(D-\tau)\,\mathrm{d}D. \tag{13}$$

Note that the simple regulation is equivalent to an FFL with $\tau = 0$.

The resulting conditions for selection of FFL over simple regulation are

$$\Phi_2 > \Phi_1, \qquad \Phi_2 > 0. \tag{14}$$

Simple regulation is selected when

$$\Phi_1 > \Phi_2, \qquad \Phi_1 > 0. \tag{15}$$

Neither circuit is selected otherwise ($\Phi_1 < 0$ and $\Phi_2 < 0$).[3] For the purpose of this comparison, the FFL is chosen to have the optimal value for τ (τ which maximizes Φ_2). These considerations map the relation between the selection of these gene circuits and the environment (specifically, relations between certain integrals of the pulse distribution).

We now consider two specific environments $P(D)$ where these conditions can be solved analytically.

The FFL is not selected in the case of exponential pulse distributions

Environments in which pulses have a constant probability per unit time to end have an exponential pulse distribution

$$P(D) = D_0^{-1}\,\mathrm{e}^{-D/D_0} \tag{16}$$

where D_0 is the mean pulse duration.

Using equations (12) and (13), we find that

$$\begin{aligned}\Phi_2 &= \int_\tau^\infty D_0^{-1}\,\mathrm{e}^{-D/D_0}\varphi(D-\tau)\,\mathrm{d}D \\ &= \mathrm{e}^{-\tau/D_0}\int_0^\infty D_0^{-1}\,\mathrm{e}^{-D/D_0}\varphi(D)\,\mathrm{d}D = \mathrm{e}^{-\tau/D_0}\Phi_1 < \Phi_1.\end{aligned} \tag{17}$$

Thus, the FFL is never selected since $\Phi_2 < \Phi_1$. Simple regulation is selected when $\Phi_1 > 0$, which occurs (using equations (12) and (6)) when the mean pulse duration is long enough $D_0 > \eta/\delta f'$. When the mean pulse duration is long enough $D_0 < \eta/\delta f'$, simple regulation is not selected because of the negative effect of the short pulses in the environment. In this case, gene Z is likely to be lost from the genome on evolutionary timescales.

Hence, the FFL is not better than simple regulation in an exponential pulse environment. In the next section, we analyze an environment where the filtering properties of the FFL can be advantageous.

[3] Using the present approach, it is easy to show that a cascade design, $X \to Y \to Z$, is never more optimal than an FFL or a simple-AND design. The reason is that the cascade shows delay after X goes off, resulting in unneeded production of Z. The FFL avoids these delays because it shows a delay only after ON steps of S_x and not OFF steps [5, 6]. Indeed, cascades are not network motifs in any known sensory transcription network [2] although they are common in developmental transcription networks [59].

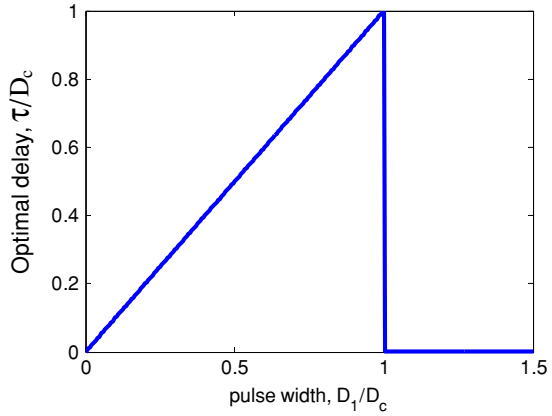

Figure 4. Optimal delay τ_0 for an FFL circuit in an environment with two types of pulses, short pulses of duration D_1 and long pulses of duration D_2. When $D_1 > D_c$, the optimal delay is $\tau_0 = 0$ and simple-AND regulation may be selected.

The FFL can be selected in bimodal distributions with long and short pulses

Consider an environment with two kinds of pulses. A pulse can have either a short duration $D_1 \ll D_c$ with probability p, or a long duration $D_2 \gg 1/\alpha$ with probability $1-p$.

The short pulses D_1 are non-beneficial, since they are shorter than the critical pulse width at which costs equal benefit, $D_1 < D_c$. In contrast, the long pulses D_2 are beneficial,

$$\varphi(D_2) = -\eta\beta D_2 + \delta f(Z_m),\ D_2 > 0. \qquad (18)$$

In this case, it is easy to calculate the optimal delay in the FFL, τ_0 (figure 4): the optimal delay is $\tau_0 = D_1$. That is, the optimal FFL has a delay, which blocks the short pulses precisely; a longer delay would reduce the benefit of the long pulses. The condition for selection of FFL over a simple-AND-gate found by solving equations (12) and (13) is that the probability of short pulses is large enough

$$p > 1 - \frac{\eta\beta}{\delta f(Z_m)}. \qquad (19)$$

The phase diagram for selection is shown in figure 5: when $\delta f(Z_m)/\eta\beta$ is small, neither circuit is selected (production costs outweigh benefits). At large $\delta f(Z_m)/\eta\beta$, the FFL is selected if short pulses are common enough (equation (19)). If short pulses are rare, simple-AND circuits are selected. At a given p, the higher the ratio of benefit to cost, $\delta f(Z_m)/\eta\beta$, the more likely the selection of simple-AND circuits.

Similar considerations apply in general to $P(D)$ with multiple peaks. Long-tailed pulse distributions, such as $P(D) \sim D^{-\gamma}$ with $\gamma > 2$, tend to show FFL selection (data not shown). Equations (12) and (13) can be used to test any distribution for its selection properties, and to generate a selection 'phase diagram' similar to figure 5.

The present model is a simplified treatment of the dynamics of these gene circuits. In the appendix, we present a more detailed model which takes into account the reactions between an enzyme and its sugar substrate, as well as graded input functions. The detailed model gives the same qualitative

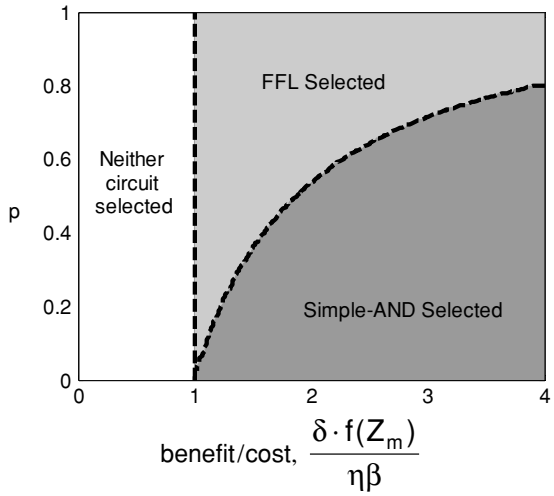

Figure 5. Selection diagram for an environment with two types of pulses, a short pulse D_1 with probability p, and a long pulse with probability $1-p$. The parameter $\delta f(Z_m)/\eta\beta$ is the ratio of benefit to production costs of protein Z. Three selection phases are shown, where FFL, simple-AND regulation or neither circuit is selected.

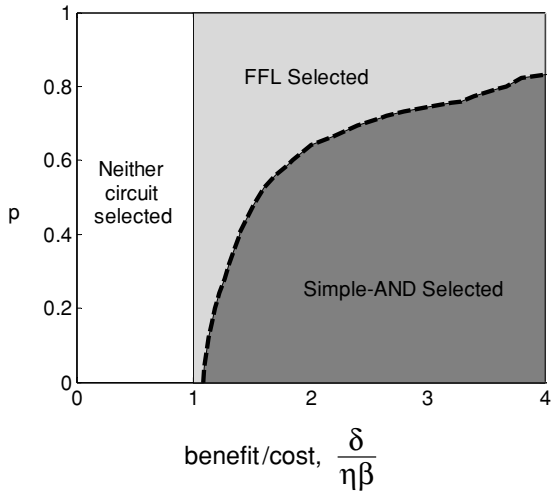

Figure 6. Selection diagram in the environment of figure 5, for the more detailed model presented in the appendix. The detailed model includes costs for Y production, graded activation of Z and $f(Z)$ based on enzyme-ligand binding. Numerical solution of the detailed model equations was used to find the optimal circuit for each value of p and $\delta f(Z_m)/\eta\beta$.

results as the analytical model discussed above (figure 6, appendix).

Discussion

We presented a simple analysis of selection of gene-regulation circuits with two inputs. This analysis is based on a cost–benefit economy in an environment with a given distribution of inputs. It yields general conditions on the environment for selection of FFLs over simple regulation circuits. We find that FFLs can be better than simple regulation in long-tailed or multi-modal environments with many short pulses.

The FFL is better when the environmental parameters are such that the cell is exposed to frequent short pulses that cannot be beneficially utilized. The FFL is not selected in environments with exponential pulse distribution. The FFL is only useful in environments where pulse duration can effectively be predicted based on whether it has outlasted a given delay. The optimal delay in the FFL can also be readily calculated for each environment.

The present cost–benefit analysis compares production costs with benefits under a time-varying environment. This cost is used as a criterion for a 'mathematically controlled comparison [25, 31, 37] between different designs. It can be extended to ask whether the optimal circuit is an evolutionarily stable solution [53]. More generally, it would be important to experimentally test whether optimality considerations are valid for gene circuits.

It is interesting to qualitatively apply the present analysis to the case of sugar systems in *E. coli*. Why is the FFL selected in some sugar system, such as the arabinose (*ara*) system [18–22], whereas simple-AND is selected in others, such as the lactose (*lac*) system [16]?

Both *ara* and *lac* systems share the same $X = \mathrm{CRP}$, a transcription activator stimulated by $S_x = \mathrm{cAMP}$, a signaling molecule produced in the cell upon glucose starvation. Thus, both *ara* and *lac* systems have the same S_x pulse distribution. According to our model, selection of circuit type would depend on the ratio of benefit to cost $\delta f(Z_\mathrm{m})/\eta\beta$, in each system. The benefit per lactose molecule (which is split into glucose + galactose) is known to be greater than the benefit per arabinose molecule (approximately 70 ATPs per lactose utilized versus approximately 30 ATPs per arabinose). Thus, the parameter $\delta f(Z_\mathrm{m})/\eta\beta$ for the *ara* system may be more to the left in figure 5 relative to the *lac* system, favoring selection of FFL in the former.

Furthermore, the availability of S_y in the natural environment of *E. coli* is different in the two systems. The sugar arabinose (S_y in the *ara* system) is thought to be far more common than lactose (S_y in the *lac* system) over most of the natural habitat of *E. coli* within its mammalian host [37]. We do not, however, know the joint probability distribution for pulses of the two signals S_x and S_y in the natural environment. The present theory suggests how differences in the joint pulse distributions of the two sugars might affect FFL selection.

Evolutionary cost–benefit analysis can also explain the selection of the values of the biochemical parameters in a given circuit [12, 25–48], as demonstrated by calculating the optimal FFL delay τ (equations (12)–(16)) as function of the environment. The value of τ is predicted to be on the time-scale of the deleterious short S_x pulses in the environment. In the *ara* system of *E. coli*, τ was experimentally found to be about 0.2 cell generations (about 20 min) [6]. Indeed, S_x (cAMP) is known to have spike-like pulses on a similar time-scale when *E. coli* cells make transitions between carbon sources [16] or undergo sudden changes in growth rate [54]. Therefore, the FFL in this system may have 'learned' the typical timescale of deleterious input pulses in the environment.

Conclusions and outlook

The present study examined the selection of a network motif, the feed-forward loop, over simpler regulation circuits, using cost–benefit analysis. The selection between simple regulation or FFL was determined as a function of the dynamic distribution of input signals in the organisms' environment.

This study makes predictions that are, in principle, experimentally testable. For example, the theory could be tested by studying a gene-regulation system in cells evolving under laboratory environments [55–58] of pulse distributions. One could then track the evolution of circuit architectures that according to the theory should be either selected or lost.

We currently have more information about the structure of some gene circuits than about the precise ecology in which they evolved. The present approach makes predictions on the environment based on the observed gene-regulation networks. It may be considered as a form of 'inverse ecology', suggesting constraints on the possible environments that could give rise to observed circuits. It would be interesting to analyze the environmental selection of the structure and parameters of other gene circuits.

Acknowledgments

We thank all members of our lab and M Savageau, M Elowitz, R Heinrich and E Klipp for discussions. This study was supported by NIH, Minerva and ISF. ED was supported by a Clore postdoctoral fellowship.

Appendix

We analyze a detailed model based on *E. coli* sugar utilization systems. The analysis employs the large separation of timescales in the problem: sugars bind and activate transcription factors within milliseconds, transcription factors bind to promoters within seconds and transcription changes protein levels on the scale of minutes or more. Rapid reactions are therefore taken at steady state within the equations for slower reaction.

Protein production dynamics

X is present at a constant level X_st. Y is regulated by X in the FFL configuration, and is not regulated in the case of simple-AND configuration. Transcription factor X becomes active when it binds S_x. When no S_x is present, X is in its inactive form, $X^* = 0$. When saturating S_x is added, $X^* = X_\mathrm{st}$. The active protein X^* binds its site in the promoter of Y with dissociation constant K_{xy}, resulting in a Michaelis–Menten term for the promoter activity of Y. As a result, Y is produced and degraded/diluted according to

$$\frac{\mathrm{d}Y}{\mathrm{d}t} = \beta_y \frac{X^*}{X^* + K_{xy}} - \alpha Y. \tag{A.1}$$

When S_y is present at saturating levels, we have

$$Y^* = Y. \tag{A.2}$$

Z is regulated by both X and Y, which bind the Z promoter with dissociation constants K_{xz} and K_{yz}, respectively. We assume for simplicity that they bind independently. Therefore, the probability that both X^* and Y^* bind their sites in the Z promoter is the product of the Michaelis–Menten probabilities of bindings,

$$P = \frac{X^*}{X^* + K_{xz}} \frac{Y^*}{Y^* + K_{yz}}. \tag{A.3}$$

The resulting dynamics of Z expression is

$$\frac{\mathrm{d}Z}{\mathrm{d}t} = \beta_z \frac{X^*}{X^* + K_{xz}} \frac{Y^*}{Y^* + K_{yz}} - \alpha Z. \tag{A.4}$$

In the FFL configuration, Y begins to be produced when X binds to S_x at time $t = 0$ according to equation (A.1):

$$Y(t) = \frac{\beta_y}{\alpha} \frac{X_{\mathrm{st}}}{K_{xy} + X_{\mathrm{st}}} (1 - \mathrm{e}^{-\alpha t}) + Y_0 \, \mathrm{e}^{-\alpha t} \tag{A.5}$$

and the analytical solution for $Z(t)$ is (equation (A.4))

$$Z(t) = \frac{a_1}{1 + a_2} \left[1 - \mathrm{e}^{-\alpha t} \left[\frac{\ln(\mathrm{e}^{\alpha t}(1 + a_2) - a_2)}{1 + a_2} + 1 \right] \right] + Z_0 \, \mathrm{e}^{-\alpha t}, \tag{A.6}$$

where

$$a_1 = \frac{\beta_z \beta_y}{K_{yz} \alpha^2} \frac{X_{\mathrm{st}}}{K_{xz} + X_{\mathrm{st}}} \frac{X_{\mathrm{st}}}{K_{xy} + X_{\mathrm{st}}} \tag{A.7}$$

$$a_2 = \frac{\beta_y}{K_{yz} \alpha} \frac{X_{\mathrm{st}}}{K_{xy} + X_{\mathrm{st}}} \tag{A.8}$$

and Y_0, Z_0 are the values of Y, Z at time $t = 0$. In a similar manner, one can readily construct the solution for environmental conditions, that changes between piecewise constant values of S_x and S_y.

Cost–benefit analysis

We now describe the effective optimization goal in order to compare the different circuits. The goal is to optimize the mean growth rate integrated over time. The growth rate is

$$g = -\eta_x \beta_x - \eta_y \beta_y \frac{X_{st}}{X_{st} + K_{xy}} - \eta_z \beta_z \frac{X_{st}}{X_{st} + K_{xz}} \frac{Y}{Y + K_{yz}} + \delta[ZS_y], \tag{A.9}$$

where $\eta_x \beta_x$, $\eta_y \beta_y$ and $\eta_z \beta_z$ are the growth cost for producing X, Y and Z. The last term represents the benefit from S_y metabolism, which is proportional to the action of enzyme Z and its substrate S_y upon binding. Z and S_y form a complex $[ZS_y]$ whose concentration at equilibrium is

$$[ZS_y] = K_z [Z][S_y] \tag{A.10}$$

where K_z is the dissociation constant of enzyme Z to sugar S_y. Two conservation laws for Z and S_y hold

$$[Z_{\mathrm{T}}] = [Z] + [ZS_y], \tag{A.11}$$

$$\left[S_y^{\mathrm{T}}\right] = [S_y] + [ZS_y], \tag{A.12}$$

where Z_{T} and S_y^{T} are the total (bound and unbound) concentrations of Z and S_y.

Equations (A.10)–(A.12) can be solved to yield a quadratic form for $[ZS_y]$:

$$[ZS_y] = \frac{[S_y^{\mathrm{T}}] + [Z^{\mathrm{T}}] + K_z - \sqrt{([S_y^{\mathrm{T}}] - [Z^{\mathrm{T}}])^2 + 2K_z([S_y^{\mathrm{T}}] + [Z^{\mathrm{T}}]) + K_z^2}}{2}. \tag{A.13}$$

Note equation (A.13) for $[ZS_y]$ reduces to standard Michaelis–Menten forms when enzyme concentration is much lower than the sugar concentration, or vice versa.

The benefit function is the rate of metabolism of S_y times the growth advantage per S_y molecule metabolized, δ_0. In the Michaelis–Menten enzyme picture, the velocity of enzyme Z is $v[ZS_y]$, and

$$\delta f(Z) = \delta_0 v [ZS_y], \tag{A.14}$$

where v is the velocity of enzyme Z.

Optimal designs

We compare the FFL and the simple-AND circuits under different environmental conditions. For a given environmental conditions ($S_x(t)$ and $S_y(t)$ profiles), we calculated the dynamics of Y and Z using equations (A.1)–(A.14). Then, using the fitness function (equation (A.9)), we calculated the temporally integrated growth rate of cells with FFL or simple-AND circuits. We optimized the growth rate of cells by finding the optimal values for the affinity and production constants $\beta_y, \beta_z, K_{xy}, K_{yz}, K_{xz}$ that give maximal growth. The optimization was done separately for the FFL and for the simple-AND configurations by using numerical Nelder–Mead simplex optimization (Matlab 6.5). The optimal growth rate of the two circuits was used to calculate the selection diagram (figure 6).

Glossary

Cost–benefit analysis. Evolutionary analysis that is based on the cells economy of costs and benefits. Production of proteins costs energy and other resources and therefore reduces the cells growth rate. The benefit comes from the function of the proteins (for example, the utilization of sugar by enzymes) that increases the growth rate. Cost–benefit analysis can design the optimal protein levels that maximize a fitness function such as growth rate.

Inverse ecology. Finding constraints on the possible ecology of an organism based on the structure of the gene circuits that have evolved in that ecology.

Phase diagram. Diagram that sections a space, whose axes are parameters of the system, into regions in which the system behavior has a particular characteristic. When the region boundaries are crossed, the system characteristic abruptly changes.

Simple regulation, simple-AND-gate regulation. A configuration where transcription factor X and transcription factor Y both regulate gene Z, but X does not regulate Y and vice versa. Both inputs are needed to be active in order to cause transcription of the gene.

Feed-forward loop (FFL). A gene circuit in which transcription factor *X* regulates transcription factor *Y* and both regulate gene *Z*. In this study, we considered an FFL where both *X* and *Y* are needed to activate *Z* (an AND-gate coherent type-1 FFL according to [5]).

References

[1] Shen-Orr S S, Milo R, Mangan S and Alon U 2002 *Nat. Genet.* **31** 64–8
[2] Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D and Alon U 2002 *Science* **298** 824–7
[3] Milo R, Itzkovitz S, Kashtan N, Levitt R, Shen-Orr S, Ayzenshtat I, Sheffer M and Alon U 2004 *Science* **303** 1538–42
[4] Alon U 2003 *Science* **301** 1866–7
[5] Mangan S and Alon U 2003 *Proc. Natl Acad. Sci. USA* **100** 11980–85
[6] Mangan S, Zaslaver A and Alon U 2003 *J. Mol. Biol.* **334** 197–204
[7] Eichenberger P *et al* 2004 *PLoS Biol.* **2** 1664–83
[8] Lee T I *et al* 2002 *Science* **298** 799–804
[9] Teichmann S A and Babu M M 2004 *Nat. Genet.* **36** 492–6
[10] Davidson E H 2002 *Science* **295** 1669–78
[11] Odom D T *et al* 2004 *Science* **303** 1378–81
[12] Conant G C and Wagner A 2003 *Nat. Genet.* **34** 264–6
[13] Buchler N E, Gerland U and Hwa T 2003 *Proc. Natl Acad. Sci. USA* **100** 5136–41
[14] Setty Y, Mayo A E, Surette M G and Alon U 2003 *Proc. Natl Acad. Sci. USA* **100** 7702–7
[15] Bolouri H and Davidson E H 2002 *Bioessays* **24** 1118–29
[16] Kremling A, Bettenbrock K, Laube B, Jahreis K, Lengeler J W and Gilles E D 2001 *Metab. Eng.* **3** 362–79
[17] Ozbudak E M, Thattai M, Lim H N, Shraiman B I and Van Oudenaarden A 2004 *Nature* **427** 737–40
[18] Wilcox G, Meuris P, Bass R and Englesbe E 1974 *J. Biol. Chem.* **249** 2946–52
[19] Casadaban M J 1976 *J. Mol. Biol.* **104** 557–66
[20] Johnson C M and Schleif R F 1995 *J. Bacteriol.* **177** 3438–42
[21] Schleif R 1996 Two positively regulated systems, ara and mal *Escherichia coli and Salmonella* vol 1 ed F C Neidhardt (Washington, DC: ASM Press) pp 1300–9
[22] Schleif R 2000 *Trends Genet.* **16** 559–65
[23] Brown C T and Callan C G Jr 2004 *Proc. Natl Acad. Sci. USA* **101** 2404–09
[24] Rosen R 1967 *Optimality Principles in Biology* (London: Butterworths)
[25] Savageau M 1976 *Biochemical Systems Analysis: A Study of Function and Design in Molecular Biology* (Reading, MA: Addison-Wesley)
[26] Heinrich R and Holzhutter H G 1985 *Biomed. Biochim. Acta* **44** 959–69
[27] Arkin A and Ross J 1994 *Biophys. J.* **67** 560–78
[28] Melendez-Hevia E, Waddell T G and Montero F 1994 *J. Theor. Biol.* **166** 201–20
[29] Heinrich R and Klipp E 1996 *J. Theor. Biol.* **182** 243–52
[30] Heinrich R and Schuster S 1996 *The Regulation of Cellular Systems* (London: Chapman and Hall)
[31] Savageau M A 1998 *Genetics* **149** 1677–91
[32] Klipp E and Heinrich R 1999 *Biosystems* **54** 1–14
[33] McAdams H H and Arkin A 2000 *Curr. Biol.* **10** R318–20
[34] Ouzounis C A and Karp P D 2000 *Genome Res.* **10** 568–76
[35] Timothy S G, Cantor C R and Collins J J 2000 *Nature* **403** 339–43
[36] Becskei A and Serrano L 2000 *Nature* **405** 590–3
[37] Savageau M A 2001 *Chaos* **11** 142–59
[38] Ibarra R U, Edwards J S and Palsson B O 2002 *Nature* **420** 186–9
[39] Hasty J 2002 *Proc. Natl Acad. Sci. USA* **99** 16516–8
[40] Yokobayashi Y, Weiss R and Arnold F H 2002 *Proc. Natl Acad. Sci. USA* **99** 16587–91
[41] Lenski R E, Ofria C, Pennock R T and Adami C 2003 *Nature* **423** 139–44
[42] Fong S S, Marciniak J Y and Palsson B O 2003 *J. Bacteriol.* **185** 6400–8
[43] Xie G, Keyhani N O, Bonner C A and Jensen R A 2003 *Microbiol. Mol. Biol. Rev.* **67** 303–42
[44] Thattai M and Shraiman B I 2003 *Biophys. J.* **85** 744–54
[45] McAdams H H, Srinivasan B and Arkin A P 2004 *Nat. Rev. Genet.* **5** 169–78
[46] Wall M E, Hlavacek W S and Savageau M A 2004 *Nat. Rev. Genet.* **5** 34–42
[47] Zaslaver A, Mayo A E, Rosenberg R, Bashkin P, Sberro H, Tsalyuk M, Surette M G and Alon U 2004 *Nat. Genet.* **36** 486–91
[48] Reich J G 1983 *Biomed. Biochim. Acta.* **42** 839–48
[49] Endy D, You L, Yin J and Molineux I J 2000 *Proc. Natl Acad. Sci. USA* **97** 5375–80
[50] Segre D, Vitkup D and Church G M 2002 *Proc. Natl Acad. Sci. USA* **99** 15112–7
[51] Rosenfeld N, Elowitz M B and Alon U 2002 *J. Mol. Biol.* **323** 785–93
[52] Perkins T J, Hallett M and Glass L 2004 *J. Theor. Biol.* **230** 289–99
[53] Maynard Smith J 1974 *J. Theor. Biol.* **47** 209–21
[54] Weber J, Kayser A and Rinas U 2005 *Microbiology* **151** 707–16
[55] Dykhuizen D E and Hartl D L 1983 *Microbiol. Rev.* **47** 150–68
[56] Elena S F and Lenski R E 2003 *Nat. Rev. Genet.* **4** 457–69
[57] Cooper T F, Rozen D E and Lenski R E 2003 *Proc. Natl Acad. Sci. USA* **100** 1072–7
[58] De Visser J A and Lenski R E 2002 *BMC Evol. Biol.* **2** 19
[59] Rosenfeld N and Alon U 2003 *J. Mol. Biol.* **13** 645–54