Optimal Gene Circuit Design

10.1 Introduction

We have seen that evolution converges again and again to the same network motifs. This suggests that network motifs are selected because they confer an advantage to the cells, as compared to other circuit designs. Can one develop a theory that explains which circuit design is selected under a given environment?
In this chapter, we will begin to explore how to apply the theory of natural selection to gene circuits. We will discuss the forces that can drive evolutionary selection, using bacteria as simple model system. The circuit that is selected, according to this theory, offers an optimal balance between the costs and benefits in a given environment.
Are cellular circuits really optimal? It is well known that most mutations and other changes to the cells' networks cause a decrease in performance. To more precisely understand evolutionary optimization, one needs to define a *fitness function* that is to be maximized. One difficulty in optimization theories is that we may not know the fitness function in the real world. For example, we currently do not know the fitness functions of cells in complex organisms. Such cells live within a society of other cells, the different tissues of the body, in which they play diverse roles. Fitness functions might not even be well defined in some cases; disciplines such as psychology and economics deal with processes that do not appear to optimize a single fitness function, but only "satisfice" (Simon, 1996) in the sense of fulfilling several conflicting and incomparable constraints (see chapter XX- multi-objective optimality).

My view is that optimality is an idealized assumption that is a good starting point for generating testable hypotheses on gene circuits. The idea is to understand the constraints under which a circuit might have evolved. You can end up concluding that a certain circuit is a historical accident or a messy, vaguely good enough solution, but it is a mistake to start by assuming this in advance.

This chapter will therefore treat the simplest systems in which one can describe the fundamental forces at play during natural selection. For additional examples, refer to the work on optimality in metabolic networks in books by Savageau, Heinrich and Schuster, Palsson and optimality in animals by McNeil Alexander (see Further Reading in Chapter 1).

Our first question is: what sets the expression level of a protein? Why are some proteins produced at a few copies per cell, others at thousands, and yet others at tens or hundreds of thousands?

10.2 Optimal expression level of a protein under constant conditions

We begin with a situation that is simple enough so that fitness can be precisely defined: bacteria that grow in a constant environment that is continually replenished. In this case, the **fitness F** is the growth rate of the cells. The number of cells N grows exponentially with time at rate F until they get too dense:

$$N(t) = N(0)\exp(Ft) \quad (10.2.1)$$

Now, if two species with different values of F compete for growth and utilize the same resources, the one with higher F will outgrow the other and inherit the test tube. Thus, evolutionary selection in this simple

case will tend to maximize F over time. This type of evolutionary selection process was elegantly described by G.F. Gause in *The Struggle for Existence* (Gause, 1934).

Fitness can help us address our question: What determines the level of expression of a protein? To be specific, we will consider a well-studied gene system, the *lac* system of *E. coli*, which was mentioned in previous chapters. The *lac* system encodes proteins such as LacZ, which breaks down the sugar **lactose** for use as an energy and carbon source. When fully induced, *E. coli* makes about z=60,000 copies of the LacZ protein per cell. Why not 50,000 or 70,000? What considerations determine the expression level of this protein?

Notice that we ask "*Why* the cell makes 60,000 copies" and not "*How* the cell makes 60,000 copies". "How" questions relate to mechanisms such as the regulatory system, the promoter sequence, and so on, which are very well characterized in the lac system. "Why" questions aim to place the system and its design within a wider theory, in this case optimality theory.

Optimality theory predicts that a protein expression level is selected that *maximizes the fitness function*. Here, the fitness function is the fitness as a function of the number of copies of the protein expressed in the cell, F(z) (Fig 12.1). In principle, F(z) can have local maxima and deep vallies (Fig 12.2). A journey on this fitness function can get stuck on local maxima or unpassable vallies. Random mutations can cause noise along this journey, especially when population sizes are small, an effect called genetic drift. Therefore, it is unclear a-priory whether evolution can reach the maximum and if so, how long this might take.

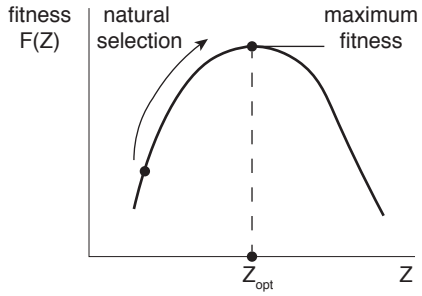

Figure 12.1

To answer such questions requires an experiment. We will consider the simplest environment possible, in which conditions are constant and do not change with time. In the case of LacZ, this means an environment with a constant concentration of the sugar lactose. The fitness is composed of two terms: the **cost** of the protein and the **benefit** it provides to the cells, both in units of growth rate, such that **F=benefit-cost**. Erez Dekel, when he was a postdoc with me, designed an experiment that measured benefit, cost and tested the predictions of optimality theory in the lac system (Dekel 2005).

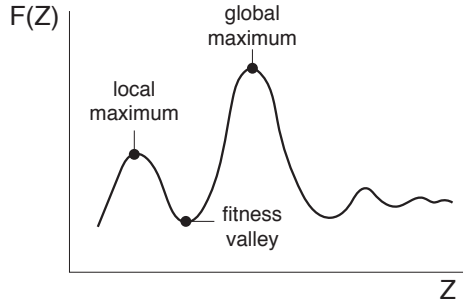

Figure 12.2

10.2.2 Cost of the LacZ protein

To experimentally measure the cost function, Erez Dekel used a classic experimental tool, the inducer IPTG. IPTG is a chemical that mimics the structure of lactose. IPTG binds to the lac repressor and causes expression of the Lac proteins, but IPTG cannot be metabolized by the cells. Thus, IPTG confers no benefit on its own, and is called a gratuitous inducer.

To measure the cost of the lac system, IPTG was used to induce the lac system to various levels in the absence of lactose. The cells grew on another carbon source, glycerol[1]. Expression of LacZ reduced the growth rate of the cells (Figure 12.3). The cost, equal to the reduction in growth rate, is found to be a *nonlinear* function of Z: the more proteins produced, the larger the cost of each additional protein. Why is the cost a nonlinearly increasing function of Z? The reason is that production and maintenance of the protein not only requires the use of the cells' resources, but also reduces the resources available to other useful proteins. To describe this, we can assume that the growth rate of the cell depends on an internal resource R (such as the amount of free ribosomes in the cell or the cells energy balance[2]). The growth rate is typically a saturating function of resources such as R, following a Michaelis function:

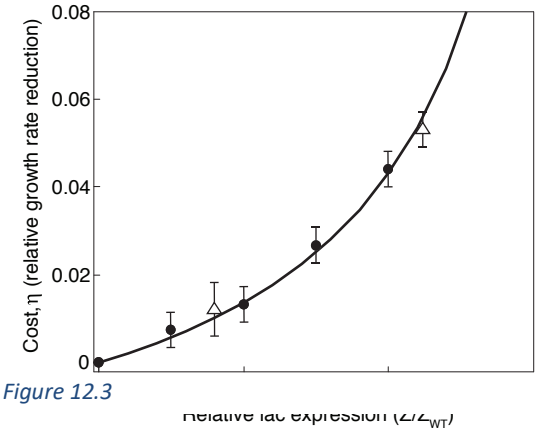

Figure 12.3

$$f \sim \frac{R}{K+R} \qquad (10.2.3)$$

The production of protein Z, and its maintenance, places a burden on the cells. This burden can be described as a reduction in the internal resource R, so that each unit of protein Z reduces the resource by a small amount, $R \to R - \epsilon z$. The upshot is that the reduction in growth rate begins to diverge when so much Z is produced that R begins to be depleted (see mathematical derivation in solved Exercise 10.4):

$$c(z) = \frac{c_0 z}{1-\frac{z}{M}} \qquad (10.2.4)$$

This cost function tells us that when only a few copies of the protein are made, the cost is approximately linear with protein level and goes as $c(z) \sim c_0 z$. The cost of a single protein is about 10^-6, which makes sense because there are about 10^6 proteins in the bacterial cell. The cost increases more steeply when z becomes comparable to an upper limit of expression, M, when it begins to seriously interfere with essential functions of the cell. Proteins cannot come too close to the point z = M, where the cost function diverges. In this experiment, M is about twice Zwt.

The experimental measurements of the cost function agree reasonably with Equation10.2.4 (Figure 12.3). They show that the relative reduction in growth rate due to the fully induced *lac* system is about 4.5%. Note that this cost of a few percent makes sense, because the fully induced Lac proteins make up a few percent of the total amount of proteins in the cell.

10.2.1 The benefit of the LacZ protein

We now turn to the benefit, defined as the increase in growth rate due to the action of the protein. In the case of LacZ, the benefit is proportional to the rate at which LacZ breaks down its substrate, lactose. The rate of the enzyme LacZ is well-described by standard Michaelis–Menten kinetics (see Appendix A). Hence, LacZ breaks down lactose at a rate that is proportional to the number of copies of the protein, z, times a saturating function of the concentration of lactose, L:

[1] Control experiments showed that IPTG itself is not toxic to the cells. For example, IPTG does not measurably affect the growth rate of cells in which the *lac* genes are deleted from the genome.

[2] The cost in this experiment is due primarily to the action of the transporter lacY. When lacY imprts a lactose molecule, it exports a proton (antiporter mechanism). This reduces the membrane potential (XX), and thus R can be the proton motive force.

$$b(z,L) = \frac{b_0 z L}{K+L} \quad (10.2.2)$$

where K is the Michaelis constant[3] and b_0 is the maximal growth rate advantage per LacZ protein at saturating lactose. Benefit therefore grows linearly with protein level z.

Benefit was measured by keeping the system maximally induced by means of IPTG, and by measuring growth rates in the presence of different levels of lactose. The observed benefit function rose with lactose levels, and was well described by Equation 10.2.2. (Fig 12.4). The relative increase in growth rate due to the fully induced level of LacZ with saturating amounts of the sugar lactose is about 17% under the conditions of the experiment.

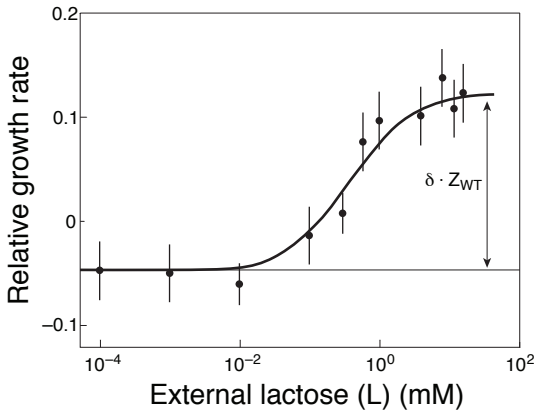

Figure 12.4

10.2.3 Fitness function and optimal expression level

Now that we have the cost and benefit functions, we can calculate the **fitness function**, equal to the difference between benefit and cost. The fitness function is the growth rate of cells that produce z copies of LacZ in an environment with a lactose concentration of L:

$$F(L,z) = benefit - cost = \frac{b_0 z\, L}{K+L} - \frac{c_0 z}{1-\frac{z}{M}} \quad (10.2.5)$$

The fitness function displays a maximum, an expression level of protein z that maximizes growth rate, as shown in Figure 12.5. The position of this maximum depends on lactose level, L. The optimal protein level z_{opt} can be found by taking the derivative of the fitness function with respect to z:

$$d F / d z = 0 \quad (10.2.6)$$

Showing that the optimal expression level increases with lactose L

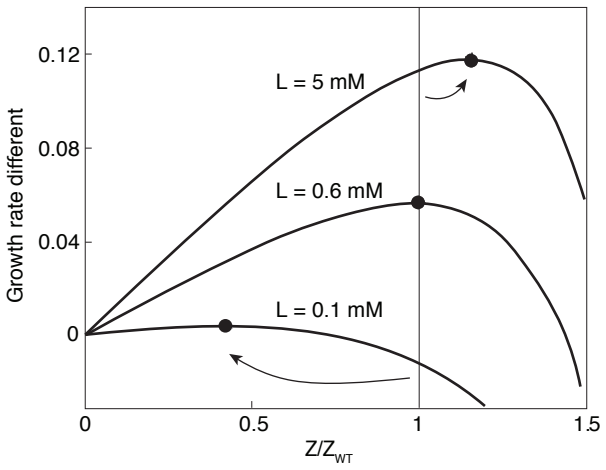

Figure 212.5

$$z_{opt} = M\left(1 - \sqrt{\frac{c_0(K+L)}{b_0\, L}}\right) \quad (10.2.7)$$

[3] Here we Z denotes LacZ and LacY proteins- LacY is the lactose transporter which is made at levels proportional to lacZ because both are on the same operon. The Michaelis constant in this case is that of the transporter LacY, K = 0.4 m*M*. This is because the influx rate of lactose is limiting under most conditions. The experiment included glycerol as a backup carbon source, so that cells can grow without lactose.

because lactose increases the benefit per LacZ enzyme, and hence increases the selection pressure to produce more enzymes. The fully induced wild-type expression level, z_{WT} is predicted to be optimal when L ~ 0.6 m*M* under these experimental conditions, as shown in Figure 12.5. Growth at higher lactose levels is predicted to be maximized by higher levels than the wild-type level of about z_{WT} = 60,000/cell. Conversely, low levels of lactose are predicted to have lower optimal expression levels (Figure 12.5).

When there is no lactose in the environment, the optimal level is $Z_{opt} = 0$, because proteins confer only costs and no benefits. In fact, zero expression is optimal as long as lactose L is lower than a threshold Lc, because costs exceed benefits. The threshold L_c can be found by asking when z_{opt} in Equation 10.2.7 becomes equal to zero:

$$z_{opt} = 0 \; when \; L < L_c = K\left(\frac{b_0}{c_0} - 1\right)^{-1} \qquad (10.2.8)$$

In the conditions of the experiments described above, the threshold level of lactose needed for selection of the gene system is $L_c \sim 0.05$ m*M*. If lactose environments with $L<L_c$ persist for many generations, the organism will tend to lose the gene encoding LacZ. The loss of unused genes is a well-known phenomenon; for example, bacteria grown in a chemostat[4] on glucose medium with no lactose lose the *lac* genes within a few days (Hartl and Dykhuizen, 1984).

10.2.4 Cells reach optimal LacZ levels in a few hundred generations in laboratory evolution experiments

We tested the predictions of this cost–benefit analysis by a **laboratory evolution experiment.** The evolution experiment used the technique of **serial dilution.** *E. coli* cells were grown in tubes with a specified level of lactose, L. Every day, 1/100 of the cells from each tube were passed to a tube with fresh medium (Fig 12.6). The cells grew in the tube until they reached stationary phase. The next morning, 1/100 of the cells were again passed to a fresh tube, and so on. Thus, every day, the cells grew 100-fold, corresponding to $\log_2(100) = 6.6$ generations. Richard Lenski has been evolving *E. coli* for a few decades with serial dilution, reaching tens of thousands of generations (Good et al, 2107).We ran the experiment for several months, with seven tubes in parallel, each with a different lactose level, L=0,0.1,0.2,0.5,1, 2and 5 mM. IPTG was added to the tubes, to make sure that lac system is fully induced. The concentration of the LacZ protein was monitored over time.

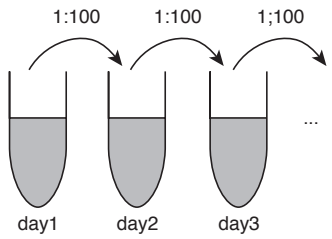

Figure 12.6

The cells heritably changed their LacZ expression level. The LacZ protein level reached the predicted optimal level within several hundred generations (Figure 12.7 and Figure 12.8). Cells growing with no lactose lost their lac expression altogether. These cells could no longer grow on lactose as the sole nutrient, (the experiments had a backup nutrient, glycerol). Cells growing on 0.1 and 0.2mM lactose reached levels of expression lower than wild-type. Cells growing with 0.5mM lactose kept close to the wild-type expression level of 60,000 per cell, as predicted.

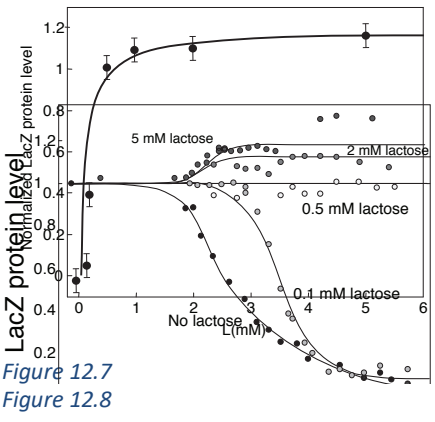

Figure 12.7
Figure 12.8

[4] A chemostat is a device that keeps bacteria growing at a constant growth rate, by supplying a constant flow of fresh medium into a mixed aerated chamber, from which medium with cells is removed at the same rate. Cell generation time is locked onto the time for exchange of half of the medium in the chamber (Novick and Weiner, 1957; Balagadde et al., 2005; Ronen and Botstein, 2006.)

Cells growing with more than 0.5mM lactose reached higher levels of expression. Cells evolving at the highest lactose level, 5mM, reached the predicted expression of about 20% more lacZ. Then at about 450 generations, they showed an unexpected jump to even higher levels. This is the great thing about evolution experiments- they often surprise you.

Analysis of the evolutionary dynamics indicated that the cells reached their optimal, adapted levels in each case by means of a mutation that changed the LacZ protein level (exc XX). For each lactose concentration, there are on the order of 100 possible mutations that can reach the desired optimal expression level. At zero lactose, there is the equivalent of 1000 mutations that lose expression altogether. Many of these mutants arise in parallel in each tube and outgrow the wild-type cells, eventually taking over the tube. For the selection pressures in this experiment, on the order of a few percent, this takes on the order of hundreds of generations.

In summary, proteins like LacZ have cost and benefit, which can be used to calculate a fitness function. This fitness function can predicts the optimal protein level in each environment. Cells rapidly evolve to this optimal value in evolutionary experiments. This gives us a sense of the speed and precision in which biological networks can adjust parameters such as protein expression levels.

What happens when conditions change with time? We will next treat the principal way that cells deal with changes: gene regulation.

10.3 To regulate or not to regulate? Optimal regulation in changing environments

In this section, we ask why are some genes regulated and other genes expressed continually without regulation. When does it pay to regulate a gene?

Consider a variable environment. Suppose that our gene product Z provides benefit to the cells only in environmental condition C_z. For example, a sugar metabolism enzyme Z is beneficial only when the sugar is available in the environment. The environment displays condition C_z with probability p, and other conditions, in which Z is superfluous, with probability 1 – p. This probability p is called the **demand** for Z.

To analyze the optimal strategy, we compare three organisms with different designs for Z regulation. In organism one, protein Z is not regulated and is produced at a constant rate under all conditions. This is known as **constitutive expression**. In the second organism, a regulatory system R is in place, so that Z is produced only when needed, in condition C_z. Regulation has a cost: the cost of production and maintenance of the regulatory system that can read the environment, and then calculate and implement the required changes in Z production. Organism three has neither the gene for Z nor the genes for its regulation system R on its genome. It cannot express protein Z at all.

The unregulated organism constantly produces Z, with a cost c, but gains its benefit b only a fraction p of the time, when Z is in demand, so that it has a fitness

$$f_1 = p\, b - c \quad (10.3.1)$$

The second organism uses a dedicated regulatory system to produce Z only under the proper conditions.

This organism thus saves unneeded production and pays the cost, c, only a fraction p of the time.

However, it bears the cost of the regulatory system, r, all of the time:

$$f_2 = p\, b - p\, c - r \quad (10.3.2)$$

Finally, the third organism that lacks the system altogether will have fitness zero, the baseline fitness without cost or benefit of Z:

$$f_3 = 0 \qquad (10.3.3)$$

Regulation will be selected when organism two has the highest fitness, $f_2 > f_1$ and $f_2 > f_3$. This leads to the following inequalities:

$$p < 1 - r/c \text{ and } p > r/(b - c) \qquad \textit{regulation selected(10.3.4)}$$

Similarly, the unregulated design in which Z is constitutively expressed will be selected when $f_1 > f_2, f_3$, leading to the inequalities:

$$p > c/b \text{ and } p > 1 - r/c \quad (10.3.5)$$

These inequalities (Equations 10.3.4 and 10.3.5) link a property of the environment, the fraction of time p that condition C_z occurs, to the cost and benefit parameters of protein Z and its regulatory system.

The range of environments in which each of the three designs is optimal is shown in Figure 12.9. Regulation is selected at an intermediate range of demand, p. High demand tends to favor systems that are continually expressed. Constitutive expression of Z is always optimal when p = 1 (assuming b>c), because if Z is always needed, regulation becomes superfluous. When p = 0, the protein is never needed and the optimal mechanism is to never express it. Thus, in constant environments (p=0 or p=1), there is no regulation, and genes are lost if cost exceeds benefit.

There exist organisms in nature whose environment is quite constant. For example, a close cousin of *E. coli*, a bacterium called *Buchnera*, lives in symbiosis inside termites helping them digest their food. The termite hosts supply Buchnera with nutrients and stable conditions (Moran, 2002; Wernegreen, 2002; Moran, 2003; Wilcox et al., 2003). In such constant environments, every protein has either p = 1 or p = 0. These organisms indeed lose virtually all of their regulation systems, such as transcription factors. They also lose most of their genes (keeping only about 800 out of 4500 in E coli). They hold a small set of genes continually expressed. This agrees with the behavior shown in Figure 12.9, on the lines p =1 and p = 0.

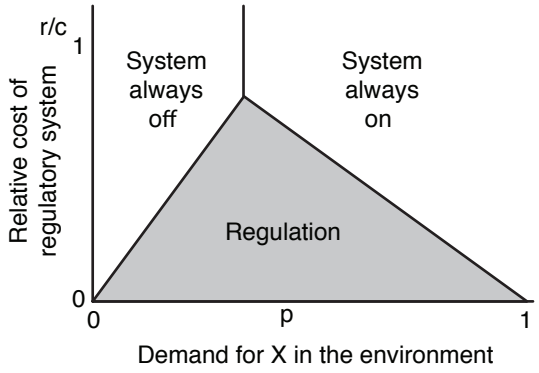

Figure 12.9

At the other extreme are bacteria that live in constantly changing and challenging environments such as the soil. These organisms have comparatively large genomes dense with prolific regulation systems.[5] These bacteria probably have 0 < p < 1 for most genes, so that regulation is selected as shown in Figure 12.9.

[5] The number of transcription factors tends to increase with the number of genes in the genome. The number of transcription factors increases as N^a, where N is the number of genes and a ~ 2 in bacteria and a ~ 1.3 in eukaryotes (Huynen and van Nimwegen, 1998; van Nimwegen, 2003). Thus, increasing the number of genes seems to require increasingly elaborate regulation mechanisms with more transcription factors per gene.[Sergei Maslov].

This analysis assumes that periods Cz in which Z is in demand are long compared to a cell generation, so that we can ignore transients in which Z levels rise when the gene is turned on. When environments change rapidly enough, a fourth strategy can be optimal – stochastic gene expression in which a fraction q of the cells express Z constitutively and the rest do not express Z. The cells gamble. This strategy is called **bet hedging**, because if Z is in demand, the cells that happen to produce Z win, and if Z is not in demand, the other cells win. The optimal fraction of Z-expressing cells, q, rises with the fraction of time that Z is in demand, p. An analysis of bet hedging is provided in solved exercise XX.

In summary, regulation makes sense if the environment is sufficiently variable.

Cost–benefit analysis gives us a way to understand the forces that drive evolutionary processes. As a final example, let's turn to cost–benefit analysis of a gene circuit, the feed-forward loop network motif.

10.4 Environmental selection of the feedforward-loop network motif

We will now try to understand, in a simplified model, under which environmental conditions this a particular motif might be selected. For motif, the purpose, we examine a common network we saw in coherent feed-forward loop (cFFL). As dynamical Chapter 3, the cFFL can perform a basic respond only function: it can filter out brief input pulses, and widespread to persistent stimuli. Although the FFL is gene is in transcription networks, not every about 40% included in an FFL. In the *E. coli* inputs are transcription network, for example, regulated of the known genes regulated by two FFL (both regulated by an FFL, and 60% are 12.10). It is by a simple two-input design which not an is selected types of circuits are shown in Figure therefore interesting to ask why the FFL in some systems and not in others.

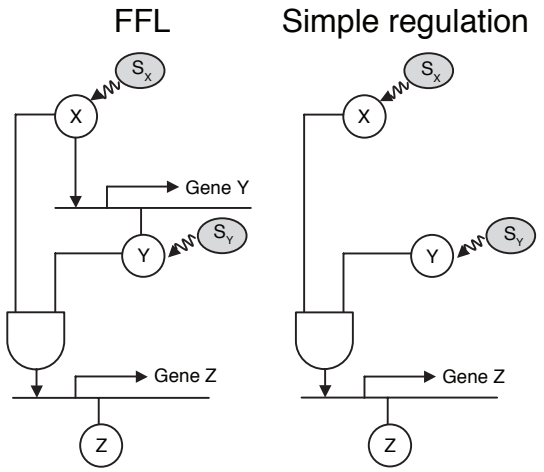

Figure 12.10

To answer this question, we will do a simplified cost–benefit analysis for the cFFL in a given dynamically fluctuating environment (Dekel et al., 2005). Here, *environment* means the time-dependent profiles of the input signals in the natural habitat of the organism. We will find conditions that the environment must satisfy in order for the FFL to be selected over a simple-regulation circuit. We will see that the FFL can be selected in environments where the input signal comes in both long and short pulses, such that the pulse duration has a wide distribution. We will also determine the optimal values of the delay of the FFL circuit as a function of the environment.

The full calculations are given in solved exercises 10.5 to 10.9. The highlights are as follows. Suppose that the system is presented with a pulse of input S_x of duration D. The fitness function, based on the cost and benefit of protein Z, can be integrated over the pulse duration, $\varphi(D) = \int_0^D f(t)dt$. This integrated fitness shows that *short pulses of input signals have a detrimental effect on growth* (Figure 12.11): they lead to a reduction in fitness. The reason for the fitness reduction is that when the input pulses are shorter than critical pulse duration, D_c, protein Z does not have time to build up to levels in which the accumulated benefit exceeds the costs of production.

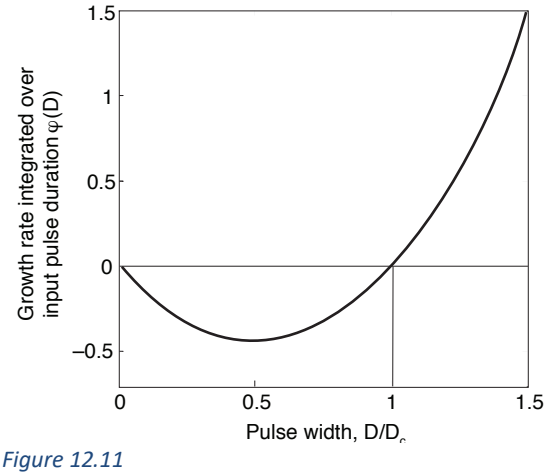

Figure 12.11

Since fitness is reduced by expression of protein Z in response to brief input pulses, a circuit that can avoid responding to brief pulses, but still allow responses to persistent pulses, can be advantageous. As we saw in Chapter 3, the coherent FFL can perform this type of filtering task. In the coherent type-1 FFL with an AND input function, Z is expressed only at a **delay** after the signals appear. Thus, only pulses of input signals longer than the delay time of the FFL will lead to Z expression.
The delay in the FFL, which we will denote T_{ON}, results from the time it takes for transcription factor Y to accumulate and cross its activation threshold for gene Z. Recall that this delay time is determined by the biochemical parameters of protein Y, namely its degradation rate, maximal level, and activation threshold for Z (Equation 3.6.5XX). The delay can therefore be tuned by natural selection to best fit the environment.

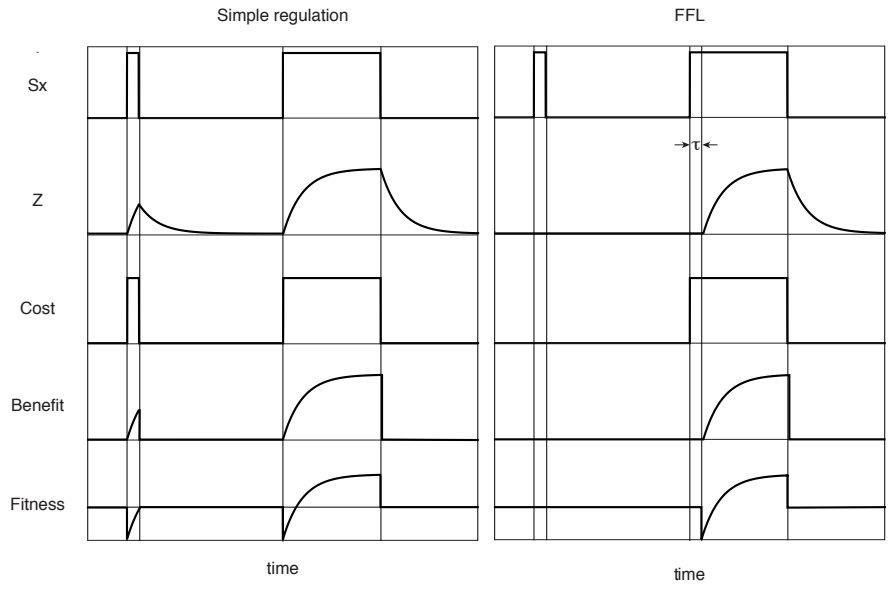

Figure 12.12

The delay in the FFL acts to filter out pulses that are shorter than T_{ON} (Figure 12.12 FFL). This avoids the reduction in growth for short pulses. However, the delay also has a disadvantage, because during long pulses, *Z is produced only at a delay and misses some of the potential benefit of the pulse* (Figure 12.12 FFL). This means that there are some situations in which the FFL does more harm than good. To assess whether the FFL confers a net advantage to the cells, relative to simple regulation, requires analysis of the full distribution of pulses in the environment, where the probability of pulse of duration D is P(D). Let us 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 average fitness, averaged over many input pulses, can be found by integrating the fitness per pulse over the pulse distribution, $F = \int P(D)\, \varphi(D)\, dD$. The design with higher average fitness has a selective advantage.

These considerations map out when each circuit has a selective advantage in terms of the environment in which they evolve. This is expressed as relations between certain integrals of the pulse distribution. Exercises 10.7 and 10.8 show that these relations can be solved exactly for certain distributions.

These solutions indicate that the FFL is selected in some environments and not in others. For example, the FFL is never selected over simple regulation in environments with an exponential pulse distribution, $P(D) \sim e^{-D/D_o}$. On the other hand, the FFL can be selected in environments with a bimodal pulse distribution, which has a probability p for short pulses that reduce fitness, and a probability 1 – p for long, beneficial pulses. The optimal delay for an FFL in such an environment is equal to the duration of the short pulses. This delay filters out the non-beneficial pulses, with minimal negative impact on fitness during long pulses.

One can draw a selection diagram that shows which circuit design has higher mean fitness (Figure 12.13). This selection diagram shows that the FFL is more fit than simple regulation in a region where brief pulses are common and the benefit-to-cost ratio of the gene system is not too high. Simple regulation is superior when brief pulses are rare. When costs exceed benefits, neither circuit is selected. Exercise 10.10 applies this to the case of two sugar systems in *E. coli*, one with simple regulation (lactose) and one with a cFFL (arabinose).

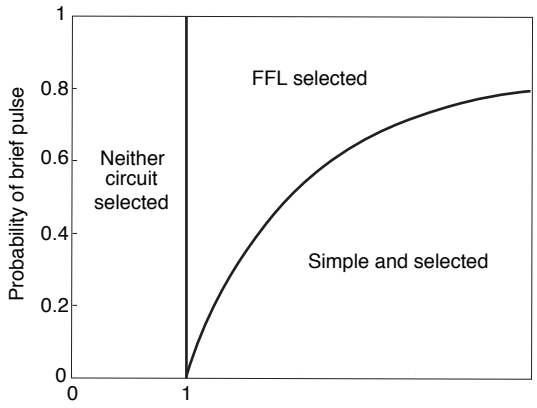

Figure 12.13

I hope that this simplified analysis gives a taste for the possibility of studying the selection of gene circuits, and their optimal parameters, in temporally changing environments.

10.5 Summary

In this chapter we discussed cost–benefit analysis as a theoretical framework for optimal circuit design. We saw that for exponentially growing bacteria, the fitness function corresponds to the cell growth rate. The cost and benefit functions can be directly measured, showing for the *lac* system a cost that increases nonlinearly with the amount of protein produced. The fitness function, equal to the difference between benefit and cost, has a well-defined optimum in each environmental condition. Optimal protein levels that maximize growth rate are reached rapidly and precisely by evolutionary selection in controlled evolutionary experiments.

We also analyzed the cost and benefit of gene regulation. We saw that gene regulation is worth maintaining only in variable environments. In constant environments, regulation tends to be lost, as is the case in organisms living as parasites within the relatively constant conditions provided by their hosts.

Finally, we saw that cost–benefit analysis can also be carried out in a dynamically changing environment, to suggest criteria for the selection of network motifs such as the coherent FFL. According to this simplified analysis, the FFL can be selected in environments that have deleterious short pulses of induction, which need to be filtered out by the function of the FFL.

We currently have more information about the structure of biological circuits than about the precise environment and ecology in which they evolved. One might imagine an inverse problem — "**inverse ecology**" — deducing information about the environment based on the observed circuits. This is based on the idea that optimal circuits contain, in a sense, *an internal model of the environment*. For example, the optimal delay time of the FFL contains information about the distributions of input pulses. Thus, an intriguing goal is to use optimality considerations to understand the molecular details of mechanisms based on the environment in which they were selected.

We will continue with these ideas in the next chapters.