6 Synthetic Biology: A Systems Engineering Perspective

Domitilla Del Vecchio and Eduardo D. Sontag

This chapter reviews some of the design challenges found in biomolecular systems from a systems engineering perspective, in particular, the problem of modularity. If components behave modularly, that is, if their behavior does not change upon interconnection, then one can predict the behavior of a circuit directly from the behavior of the composing units. In two instances of oscillating synthetic biomolecular systems, we demonstrate that, because of loading effects called "retroactivity" at interconnections, modularity does not necessarily hold in biomolecular systems. We propose a framework for quantifying retroactivity at interconnections between transcriptional circuits and present a mechanism, inspired by the design of electronic noninverting amplifiers, to counteract retroactivity.[1]

6.1 Background

Although biologists have long employed phenomenological and qualitative models to help discover the components of living systems and to describe their behaviors, the analysis of the dynamical properties of complex biomolecular reaction networks requires a more quantitative and systems-level approach. Thus, in recent years, the field of systems biology has emerged, whose focus is the quantitative analysis of cell behavior, with the goal of explicating the basic dynamic processes, feedback control loops, and signal-processing mechanisms underlying life. Complementary to systems biology is the new engineering discipline of *synthetic biology*, whose goal is to extend or modify the behavior of organisms, and to induce them to perform new tasks (Andrianantoandro et al., 2006; Endy, 2005). Through the de novo construction of simple elements and circuits, the field aims to foster an engineering framework for obtaining new cell behaviors in a predictable and reliable fashion. The ultimate goal is to develop synthetic biomolecular circuitry for a wide variety of applications from targeted drug delivery to the construction of biomolecular computers. In the process, synthetic biology helps us improve our quantitative and qualitative understanding

of biological systems through designing and constructing instances of these systems in accordance with hypothesized models. Discrepancies between expected behavior and observed behavior highlight research issues that need more study, gaps in our knowledge, or inaccurate assumptions in models. One of the fundamental building blocks employed in synthetic biology is the process of transcriptional regulation. A transcriptional network is composed of a number of genes that express proteins that then act as transcription factors for other genes. The rate at which a gene is transcribed is controlled by the *promoter*, a regulatory region of DNA that lies upstream of the coding region of the gene. RNA polymerase binds a defined site (a specific DNA sequence) on the promoter. The quality of this site specifies the transcription rate of the gene (the DNA sequence at the site determines its chemical affinity for RNA polymerase). Although RNA polymerase acts on all of the genes, each transcription factor modulates only a particular set of target genes. Transcription factors affect the transcription rate by binding specific sites on the promoter region of the regulated genes. When bound, they change the probability per unit time that RNA polymerase binds the promoter region. Transcription factors thus affect the rate at which RNA polymerase initiates transcription. A transcription factor can act as a *repressor* when it prevents RNA polymerase from binding to the promoter site. A transcription factor acts as an *activator* if it facilitates the binding of RNA polymerase to the promoter. Such interactions can be generally represented as nodes connected by directed edges. Synthetic biomolecular circuits are typically fabricated in *Escherichia coli*, by cutting and pasting together coding regions and promotors (natural and engineered) according to designed structures. Because the expression of a gene is under the control of the upstream promoter region, this technique allows the production of any desired circuit of activation and repression interactions among genes. Early examples of such circuits include an activator-repressor system that can display toggle switch or clock behavior (Atkinson et al., 2003), a loop oscillator called the "repressilator" obtained by connecting three inverters in a ring topology (Elowitz and Leibler, 2000), a toggle switch obtained by connecting two inverters in a ring fashion (Gardner et al., 2000), and an autorepressed circuit (Becskei and Serrano, 2000; figure 6.1).

Several scientific and technological developments over the past four decades have set the stage for the design and fabrication of early synthetic biomolecular circuits. An early milestone in the history of synthetic biology was the discovery in 1961 of mathematical logic in gene regulation (Jacob and Monod, 1961). Only a few years later, special enzymes that can cut double-stranded DNA at specific recognition sites, known as restriction sites, were discovered (Arber and Linn, 1969). These enzymes, called restriction enzymes, were a major enabler of recombinant DNA technology. One of the most celebrated products of such technology is the large-scale production of insulin by *E. coli* bacteria, which serve as cellular factories (Villa-Komaroff et al.,

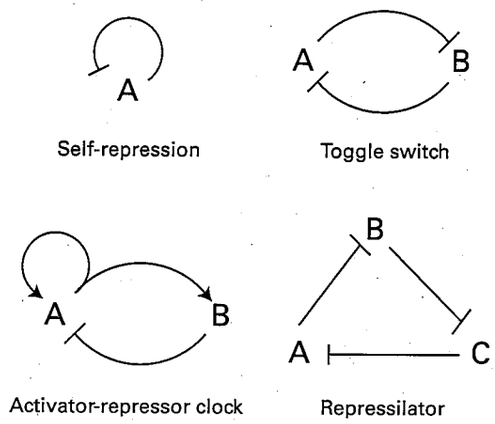

Figure 6.1
Early transcriptional circuits that have been fabricated in the bacterium *E. coli*: the self-repression circuit (Becskei and Serrano, 2000), the toggle switch (Gardner et al., 2000), the activator-repressor clock (Atkinson et al., 2003), and the repressilator (Elowitz and Leibler, 2000). Each node represents a gene and each arrow from node Z to node X indicates that the transcription factor encoded in z, denoted Z, regulates gene x. If z represses the expression of x, the interaction is represented by $Z \dashv X$. If z activates the expression of x, the interaction is represented by $Z \rightarrow X$ (Alon, 2007)

1978). The development of recombinant DNA technology along with the demonstration in 1970 that genes can be artificially synthesized, provided the ability to cut and paste natural or synthetic promoters and genes in almost any fashion on plasmids of compatible size through the cloning process (Alberts et al., 1989). The polymerase chain reaction (PCR), devised in the 1980s, allows the exponential amplification of small amounts of DNA into amounts large enough to be used for transfection and transformation in living cells (Alberts et al., 1989). Today, commercial synthesis of DNA sequences and genes has become cheaper and faster, often costing less than $1 per base pair (Baker et al., 2006). Fluorescent proteins, such as GFP and its genetic variations, allow the in vivo measurement of the amount of protein produced by any target gene, providing a readout of gene circuit behavior. Circuit design also allows for external inputs in the form of inducers, used to probe the system. Inducers act by disabling repressor proteins, thus modulating the levels of transcription.

One of the current directions of the field is to create circuitry with more complex functionalities by assembling simpler circuits, such as those in figure 6.1. This tendency reflects the history of electronics after the bipolar junction transistor (BJT) was invented in 1947. In particular, a major breakthrough occurred in 1964 with the invention of the first operational amplifier (OPAMP), which led the way to standardized modular and integrated circuit design. By comparison, synthetic biology may be moving toward a similar development of modular and integrated circuit design. This

is witnessed by several recent efforts toward formally characterizing between-module interconnection mechanisms, loading (or *impedance*) effects, and OPAMP-like devices to counteract loading problems (Del Vecchio et al., 2008b; Hartwell et al., 1999; Rubertis and Davies, 2003; Saez-Rodriguez et al., 2004, 2005; Sauro and Ingalls, 2007; Sauro and Kholodenko, 2004).

Section 6.2 describes the fundamental modeling assumption made for circuit analysis and design: *modularity*, which guarantees that building blocks maintain their behavior unchanged after interconnection. This property is fundamental for predicting the behavior of a complex system by the behavior of the composing units. Section 6.3 shows how the two synthetic oscillators of figure 6.1 can be designed assuming modular composition of their building blocks. Section 6.4 shows that modularity does not necessarily hold in transcriptional circuitry in the same manner as it occurs in many other engineering systems (Willems, 1999). Here we introduce the concept of *retroactivity* to characterize any change in the dynamics of a building block due to interconnection. We describe a procedure for quantifying retroactivity and thus for designing an interconnection so as to have low retroactivity, when possible. In section 6.5, we propose the concept of an insulation device as a system that enforces modularity by working as a buffer between a component that sends a signal and one that receives the signal. Large amplification and feedback gains are the key mechanisms for the design of insulation devices in many other engineering systems. We show that simple cycles, involving phosphorylation and dephosphorylation, which are ubiquitous in natural signal transduction systems, enjoy intrinsic insulation properties, and thus have the potential to serve as synthetic biomolecular insulation devices.

6.2 The Modularity Assumption

Each node y of a transcriptional network is usually modeled as an input-output module taking as input the concentrations of transcription factors that regulate gene y and giving as output the concentration of protein expressed by gene y, denoted Y. The transcription factors regulating y appear as inputs of the transcriptional module through their association/dissociation with the promoter site of gene y. We will denote by X the protein, by X the average protein concentration, and by x the gene expressing protein X. The internal dynamics of the transcriptional module are determined by the processes of transcription and translation, which are much slower than the dynamics of transcription factor binding (Alon, 2007). The binding of transcription factors to the promoter site reaches equilibrium in seconds, whereas transcription and translation of the target gene take minutes to hours. This time-scale separation, a key feature of transcriptional circuits, leads to the following central modeling simplification.

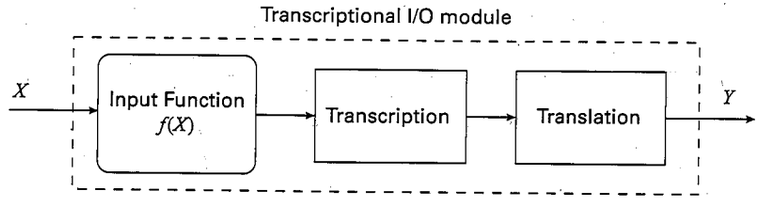

Figure 6.2
Transcriptional module modeled as an input-output system with input function given by the transcription regulation function $f(X)$ and with internal dynamics established by the transcription and translation processes.

According to the *modularity assumption*, the dynamics of transcription factor/DNA binding are considered at equilibrium, and each transcription factor concentration enters the input-output transcriptional module through a *static* input function that drives the transcription and translation dynamics (figure 6.2).

In the simplest case of one input acting as repressor or activator, the transcription regulation functions $f(X)$ take the Hill function form. When the transcriptional component takes several transcription factors as inputs, more complicated forms can be constructed from first principles (Alon, 2007).

Consider a transcriptional module with input function $f(X_1, \ldots, X_n)$. The internal dynamics of the transcriptional module usually model mRNA and protein dynamics through the processes of transcription and translation. Protein production is balanced both by decay, through *degradation*, which occurs when the protein is destroyed by specialized proteins in the cell that, for example, recognize a specific part of the protein and destroy it, and by *dilution*, which is caused by the reduction in concentration of the protein due to the increase of cell volume during growth. In a similar way, mRNA production is also balanced by degradation and dilution. Thus the dynamics of a transcriptional module are often well captured by the following ordinary differential equations:

$$\frac{dr_Y(t)}{dt} = f(X_1(t), \ldots, X_n(t)) - \alpha_1 r_Y(t), \tag{6.1a}$$

$$\frac{dY(t)}{dt} = \gamma r_Y(t) - \alpha_2 Y(t), \tag{6.1b}$$

in which r_Y denotes the concentration of mRNA translated from gene Y, the constants α_1 and α_2 indicate the mRNA and protein decay rates, and γ is a constant that establishes the rate at which the mRNA is translated.

To engineer a system with prescribed behavior, one must be able to change the physical features so as to change the values of the parameters of the model. This is

often possible. For example, the binding affinity of a transcription factor to its site on the promoter can be affected by single or multiple base pair substitutions. The protein decay rate (constant α_2 in equation [6.1b]) can be increased by adding degradation tags at the end of the gene expressing protein Y. Tags are genetic additions to the end of a sequence that modify expressed proteins in different ways such as marking the protein for faster degradation. Combinatorial promoters, which can accept multiple input transcription factors to implement regulation functions that take multiple inputs, can be realized by combining the operator sites of several simple promoters (Cox et al., 2007).

6.3 Design of Genetic Circuits under the Modularity Assumption

Based on the modeling assumptions outlined in the previous section, a number of synthetic genetic circuits have been designed and fabricated by composing transcriptional modules through input-output connection (figure 6.1). With such a procedure, one seeks to predict the behavior of a circuit from that of its composing units, once these have been well characterized in isolation. This approach is standard also in the design and fabrication of electronic circuitry.

6.3.1 The Repressilator

Elowitz and Leibler (2000) constructed the first operational oscillatory genetic circuit, which consists of three repressors arranged in ring fashion, and called it the "repressilator" (figure 6.1). The repressilator exhibits sinusoidal, limit cycle oscillations in periods of hours, slower than the period of the *E. coli* cell-division cycle. The state of the oscillator is transmitted between generations from mother to daughter cells. In the repressilator, the protein lifetimes are shortened to approximately two minutes (close to mRNA lifetimes). A dynamical model of the repressilator can be obtained by composing three transcriptional repression modules in a loop fashion:

$$\frac{dr_A(t)}{dt} = -\delta r_A(t) + f_1(C(t)), \qquad \frac{dA(t)}{dt} = r_A(t) - \delta A(t), \tag{6.2a}$$

$$\frac{dr_B(t)}{dt} = -\delta r_B(t) + f_2(A(t)), \qquad \frac{dB(t)}{dt} = r_B(t) - \delta B(t), \tag{6.2b}$$

$$\frac{dr_C(t)}{dt} = -\delta r_C(t) + f_3(B(t)), \qquad \frac{dC(t)}{dt} = r_C(t) - \delta C(t). \tag{6.2c}$$

We will consider two different cases for the input functions f_i: the symmetric case, with three identical repressions, and the nonsymmetric case, with two identical activations and one repression. For the symmetric case, we assume that

$$f_1(p) = f_2(p) = f_3(p) = \frac{\alpha^2}{1+p^n}.$$

As the regulatory functions all have negative slope, and there is an odd number of them in the loop, there is only one equilibrium. One can then invoke well-known theorems of Mallet-Paret and Smith (1990) or of Hastings et al. (1977) to conclude that, if the equilibrium point is unstable, the system admits a nonconstant periodic orbit. Thus, to obtain periodic behavior, one can search for parameter values to guarantee the instability of the equilibrium point, a procedure followed in the design of the repressilator (Elowitz and Leibler, 2000). In particular, one can show that the symmetric repressilator in equation (6.2) has a periodic solution if the parameters α, δ, and n satisfy

$$\alpha^2/\delta^2 > \sqrt[n]{\frac{4/3}{n-4/3}}\left(1+\frac{4/3}{n-4/3}\right).$$

This relationship is plotted in figure 6.3a. When n increases, the existence of an unstable equilibrium point is guaranteed for larger ranges of the other parameter values. Equivalently, for fixed values of α and δ, as n increases, the robustness of the circuit oscillatory behavior to parametric variations in the values of α and δ also increases. Of course, this "behavioral" robustness does not guarantee that other

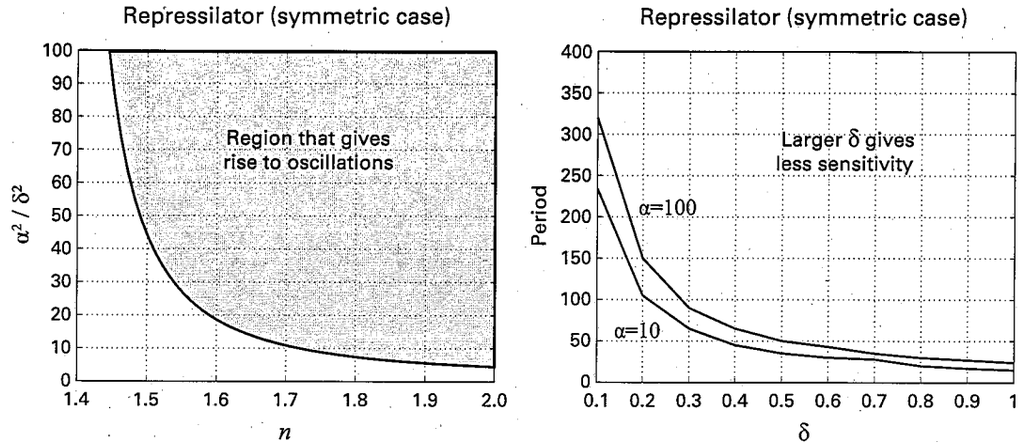

Figure 6.3
Repressilator (symmetric case). (a) Space of parameters that give rise to oscillations for the repressilator in equation (6.2). (b) Period as a function of δ and α.

important features of the oscillator, such as the period value, are insensitive to parameter variation. Numerical studies indicate that the period T approximately follows $T \propto 1/\delta$, and varies only slightly with α (figure 6.3b). From the figure, we can note that, as the value of δ increases, the sensitivity of the period to the variation of δ itself decreases. However, increasing δ would necessitate an increase of the cooperativity n. This analysis indicates a potential trade-off that should be taken into account in the design process in order to balance system complexity and the robustness of the oscillations.

A similar result for the existence of a periodic solution can be obtained for the nonsymmetric case, in which the input functions of the three transcriptional modules are modified to

$$f_1(p) = \frac{\alpha_3^2}{1+p^n}, \quad f_2(p) = \frac{\alpha^2 p^n}{1+p^n} \quad \text{and} \quad f_3(p) = \frac{\alpha^2 p^n}{1+p^n};$$

that is, two interactions are activations and one is a repression. One can verify that there is only one equilibrium point and again invoke the theorems of Mallet-Paret and Smith (1990) or Hastings et al. (1977) to conclude that if the equilibrium point is unstable, the system admits a nonconstant periodic solution. We can thus obtain the condition for oscillations again by establishing conditions on the parameters that guarantee an unstable equilibrium. These conditions are reported in figure 6.4. One can conclude that it is possible to design the circuit to be in the region of parameter space that gives rise to oscillations. It is also possible to show that, as the number of elements in the oscillatory loop increases, the value of n sufficient for oscillatory behavior decreases. The design criteria for obtaining oscillatory behavior are summarized in figures 6.3 and 6.4.

6.3.2 The Activator-Repressor Clock

Consider the activator-repressor clock diagram shown in figure 6.1, which is an example of a *relaxation oscillator*. The transcriptional module for A has an input function that takes two arguments: the activator concentration A and the repressor concentration B. The transcriptional module B has an input function that takes only the activator concentration A as its input. Let r_A and r_B represent the concentration of mRNA of activator and repressor, respectively. We consider the following four-dimensional model describing the rate of change of the species concentrations:

$$\frac{dr_A(t)}{dt} = -\delta_1/\epsilon r_A(t) + F_1(A(t), B(t)), \qquad \frac{dA(t)}{dt} = \nu(-\delta_A A(t) + k_1/\epsilon r_A(t)),$$

$$\frac{dr_B(t)}{dt} = -\delta_2/\epsilon r_B(t) + F_2(A(t)), \qquad \frac{dB(t)}{dt} = -\delta_B B(t) + k_2/\epsilon r_B(t),$$

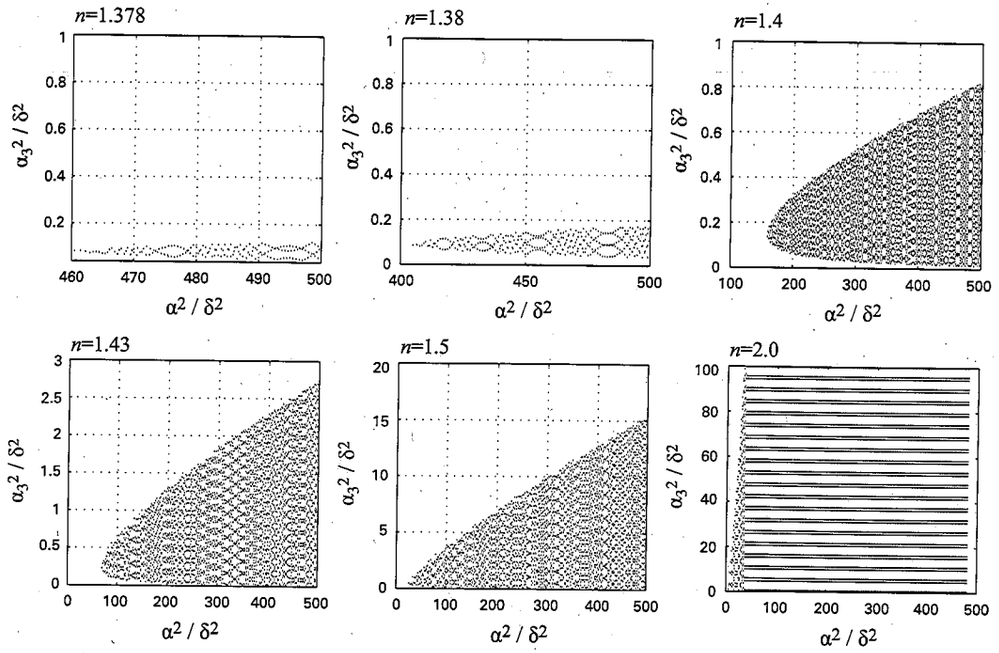

Figure 6.4
Space of parameters that give rise to oscillations for the repressilator (nonsymmetric case; El Samad et al., 2005).

in which the parameter ν regulates the difference of time scales between the repressor and the activator dynamics, and ϵ is a parameter that regulates the difference of time scales between the mRNA and the protein dynamics. The input functions F_1 and F_2 are given by

$$F_1(A,B) = \frac{K_1 A^n + K_{A0}}{1+\gamma_1 A^n + \gamma_2 B^n} \quad \text{and} \quad F_2(A) = \frac{K_2 A^n + K_{B0}}{1+\gamma_3 A^n},$$

in which K_1 and K_2 are the maximal activated transcription rates, while K_{A0} and K_{B0} are the basal transcription rates when no activator is present. The parameters $1/\gamma_i$ are the activation coefficients and are related to the affinity of the protein to the promoter site. The Hill coefficient is chosen to be $n = 2$.

The number of equilibria of the system is not influenced by the values of ϵ and of ν but is dependent on the other model parameters. The set of values of K_i, k_i, δ_i, γ_i, δ_A, and δ_B that result in the existence of a unique equilibrium can be determined by employing graphical techniques. In particular, one can plot the curves corresponding to the sets of A, B values for which $dr_B/dt = 0$ and $dB/dt = 0$ and the set of A, B

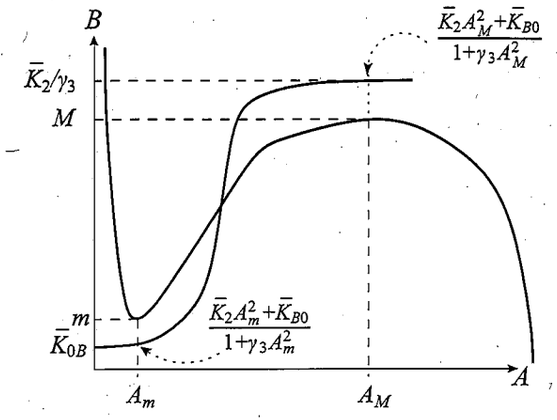

Figure 6.5
Shape of the curves in the A, B plane corresponding to $dr_B/dt = 0$, $dB/dt = 0$, and to $dr_A/dt = 0$, $dA/dt = 0$ as function of the parameters.

values for which $dr_A/dt = 0$ and $dA/dt = 0$ as in figure 6.5. The intersection of these two curves provides the equilibria of the system and conditions on the parameters can be determined that guarantee the existence of one equilibrium only.

We introduce scaled parameters:

$$\bar{K}_1 = \frac{K_1}{\delta_1/\epsilon}, \quad \bar{K}_{A0} = \frac{K_{A0}}{\delta_1/\epsilon}, \quad \bar{K}_2 = \frac{K_2}{\delta_2/\epsilon}, \quad \text{and} \quad \bar{K}_{B0} = \frac{K_{B0}}{\delta_2/\epsilon}.$$

In particular, we require that the basal activator transcription rate when B is not present, which is proportional to $\bar{K}_{A0}$, is sufficiently smaller than the maximal transcription rate of the activator, which is proportional to $\bar{K}_1$. Also, $\bar{K}_{A0}$ must be nonzero. In the case that $\bar{K}_1 \gg \bar{K}_{A0}$, one can verify that $A_M \approx \bar{K}_1/2\gamma_1$ and thus $M \approx \bar{K}_1/2\sqrt{\gamma_1\gamma_2}$. As a consequence, if $\bar{K}_1/\gamma_1$ increases, then so must $\bar{K}_2/\gamma_3$. This implies that the maximal transcription rate of the repressor divided by its protein and mRNA decay rates must be larger than the maximal transcription rate of the activator divided by its protein and mRNA decay rates. Finally, $A_m \approx 0$, and $m \approx \sqrt{\bar{K}_{A0}/\gamma_2 A_m}$. As a consequence, the smaller $\bar{K}_{A0}$ becomes, the smaller $\bar{K}_{B0}$ must be (see Del Vecchio, 2007, for details).

Given that the values of K_i, k_i, δ_i, γ_i, δ_A, and δ_B have been chosen so that there is a unique equilibrium, we numerically study the occurrence of periodic solutions as the difference in time scales between protein and mRNA, ϵ, and the difference in time scales between activator and repressor, ν, are changed. In particular, we perform bifurcation analysis with ϵ and ν, the two bifurcation parameters. These bifurcation results are summarized by figure 6.6 (see Del Vecchio, 2007, for the details of the numerical analysis).

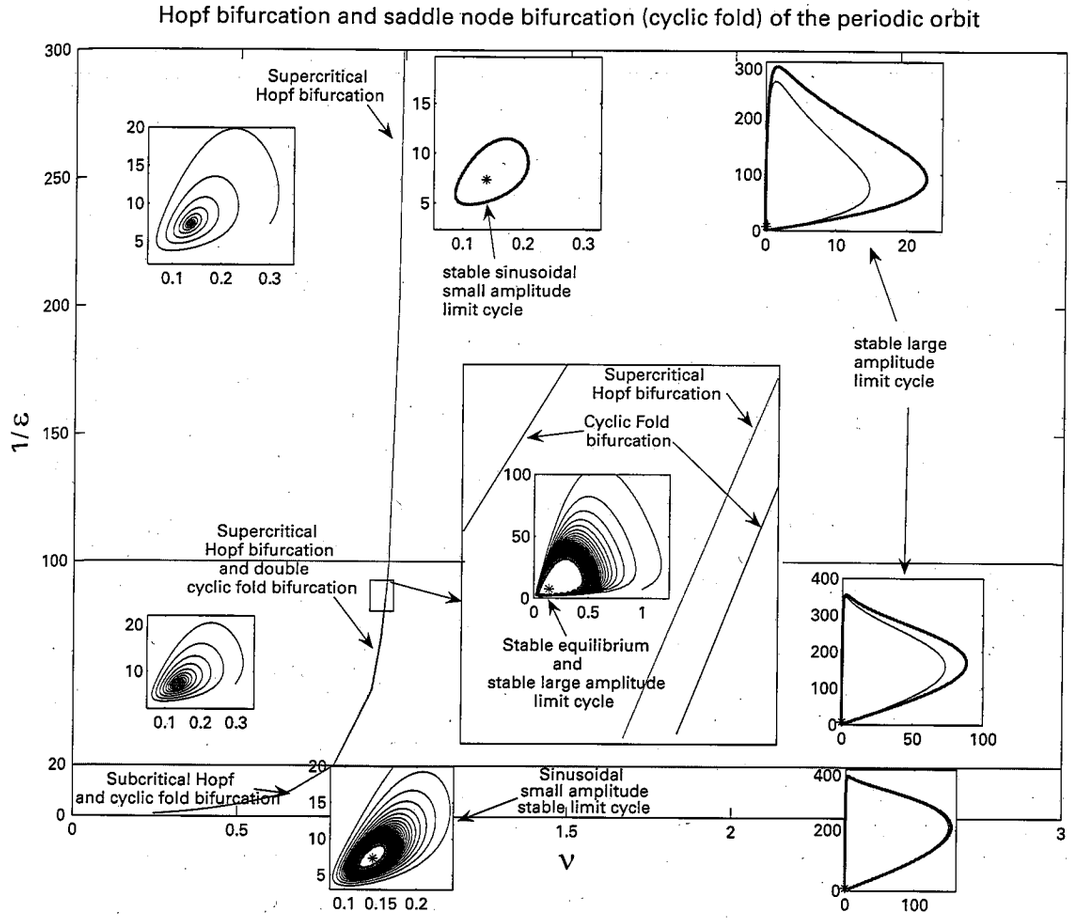

Figure 6.6
Design chart for the relaxation oscillator: For values of ν sufficiently large, one obtains sustained oscillations beyond the Hopf bifurcation, independently of the difference of time scales between the protein and the mRNA dynamics. Note also that there are values of ν for which a stable point and a stable orbit coexist and values of ν for which two stable orbits coexist. The interval of ν values for which two stable orbits coexist is too small to be able to set ν numerically in such an interval. Thus this interval is not practically relevant. The values of ν for which a stable equilibrium and a stable periodic orbit coexist, known as hard excitation, is instead relevant.

The situation described in figure 6.6 corresponds to the *hard excitation* condition (Leloup and Goldbeter, 2001) and occurs for realistic values of the separation of time scales between protein and mRNA dynamics. This simple oscillator motif described by a four-dimensional model can thus capture the features that lead to the long-term suppression of the rhythm by external inputs. *Birhythmicity* (Goldbeter, 1996) is also possible even if practically not relevant due to the numerical difficulty of moving the system to one of the two periodic orbits (see Del Vecchio, 2007; Conrad et al., 2008, for details). In terms of the ϵ and ν parameters, it is thus possible to design the system to achive oscillatory behavior: as the time scale of the activator dynamics increases with respect to the repressor dynamics, the system parameters move across a Hopf bifurcation and stable oscillations will arise. From a fabrication point of view, this can be achieved by adding suitable degradation tags to the activator protein. The region of the parameter space in which the system exhibits almost sinusoidal damped oscillations is on the left-hand side of the curve corresponding to the Hopf bifurcation. As the data of Atkinson et al. (2003) exhibit almost sinusoidal damped oscillations, it is possible that the clock is operating in a region of parameter space on the "left" of the curve corresponding to the Hopf bifurcation. If this were the case, increasing the separation of time scales between the activator and the repressor, ν, may lead to a stable limit cycle.

6.4 Beyond the Modularity Assumption: Retroactivity

The circuit design process outlined thus far relies deeply on the modularity assumption, by virtue of which the behavior of the circuit topology can be directly predicted by the properties of the composing units. For example, the monotonicity of the input functions of the transcriptional modules composing the repressilator was a key to formally showing the existence of periodic solutions. The form of the input functions in the activator-repressor clock design enabled easy predictions of the location and number of equilibria as the parameters were changed. The modularity assumption implies that, when two modules are connected to one another, their behavior does not change upon interconnection. However, a fundamental systems-engineering issue that arises when interconnecting subsystems is how the process of transmitting a signal to a downstream component affects the dynamic state of the sending component. Indeed, after designing, testing, and characterizing the input-output behavior of an individual component in isolation, it is certainly desirable if its characteristics do not change when another component is connected to its output channel. This problem, the effect of loads on the output of a system, is well understood in many fields of engineering, for example, in electrical circuit design. It has often been pointed out that similar issues arise for biological systems. Modules should have special features

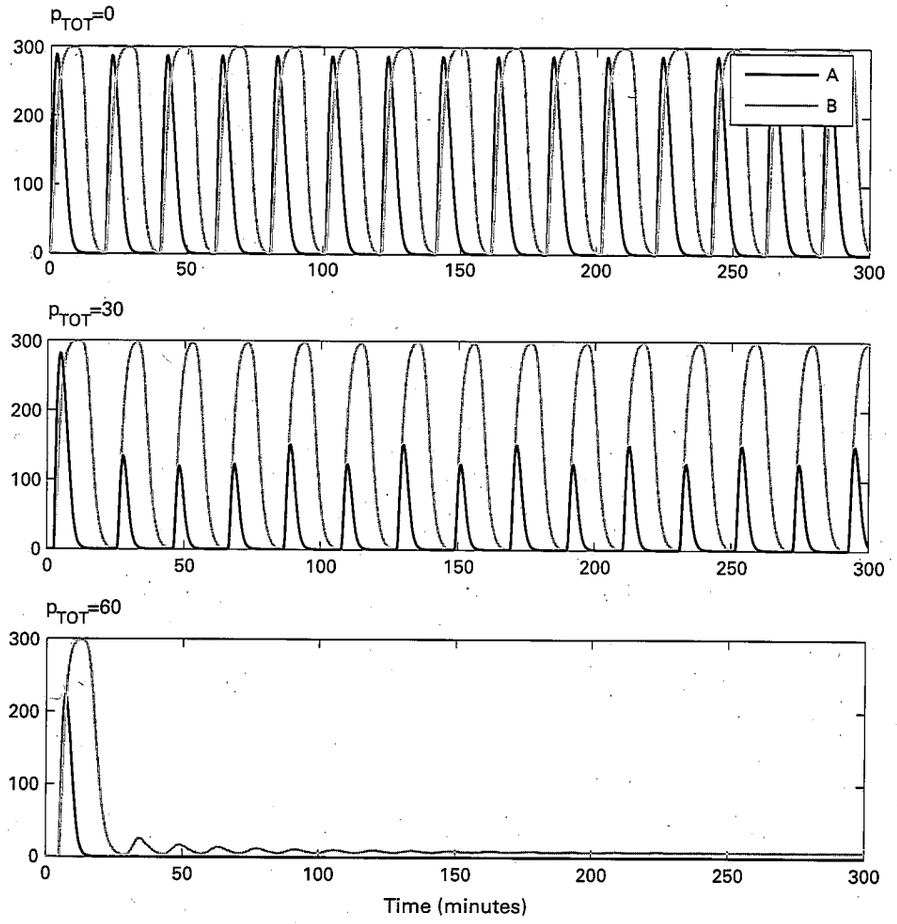

Figure 6.7
Activator-repressor clock behavior can be disrupted by a load on the activator A. As the number of downstream binding sites for A, p_{TOT}, is increased in the load, the activator and repressor dynamics lose their synchronization, and ultimately the oscillations disappear.

that allow them to be easily embedded in any system, such as zero output impedance and infinite input impedance. An extensive review on problems of loads and modularity in signaling networks can be found in Sauro (2004); Sauro and Ingalls (2007); Sauro and Kholodenko (2004), where the authors propose concrete analogies with similar problems arising in electrical circuits.

These questions are especially delicate in *synthetic* biology. For example, consider the activator-repressor clock of figure 6.1. Assume we want to employ this clock (upstream system) to drive one or more components (downstream systems), by using as its *output* signal the oscillating concentration $A(t)$ of the activator. From a systems/signals point of view, $A(t)$ becomes an *input* to the second system. The terms *upstream* and *downstream* reflect the direction in which we think of signals as traveling, *from* the clock *to* the systems being synchronized. This is only an idealization, however, because the binding and unbinding of A to promoter sites in a downstream system competes with the biochemical interactions that constitute the upstream block (retroactivity) and may therefore disrupt the operation of the clock itself (figure 6.7). One possible approach to avoid disrupting the behavior of the clock, motivated by the approach used with reporters such as GFP, is to introduce a gene coding for a new protein X, placed under the control of the same promoter as the gene for A, and to use the concentration of X, which presumably mirrors that of A, to drive the downstream system. This approach, however, has still the problem that the behavior of the X concentration in time may be altered and even disrupted by the addition of downstream systems that drain X. The net result is still that the downstream systems are not properly timed.

6.4.1 Modeling Retroactivity

We broadly call *retroactivity* the phenomenon by which the behavior of an upstream system is changed upon interconnection to a downstream system. As a simple example, consider a transcriptional component whose output is connected to downstream processes, which can be, for example, other transcriptional components (figure 6.8).

Figure 6.8
The transcriptional component takes as input u protein concentration Z and gives as output y protein concentration X.

The activity of the promoter controlling gene x depends on the amount of Z bound to the promoter. If $Z = Z(t)$, such an activity changes with time. We denote it by $k(t)$. By neglecting the mRNA dynamics, which are not relevant for the current discussion, we can write the dynamics of X as

$$\frac{dX(t)}{dt} = k(t) - \delta X(t), \tag{6.3}$$

in which δ is the decay rate of the protein. We refer to equation (6.3) as the isolated system dynamics. Now assume that X drives a downstream transcriptional module by binding to a promoter p with concentration p (figure 6.8). The reversible binding of X with p is then described by

$$\mathrm{X} + \mathrm{p} \underset{k_{\mathrm{off}}}{\overset{k_{\mathrm{on}}}{\rightleftharpoons}} \mathrm{C},$$

in which C is the protein-promoter complex and k_{on} and k_{off} are the association and dissociation rates. Because the promoter is not subject to decay, its total concentration p_{TOT} is conserved so that we can write $p + C = p_{\mathrm{TOT}}$. The new dynamics of X are therefore governed by the equations

$$\frac{dX(t)}{dt} = k(t) - \delta X(t) + \boxed{k_{\mathrm{off}} C(t) - k_{\mathrm{on}}(p_{\mathrm{TOT}} - C(t))X(t)}, \tag{6.4a}$$

$$\frac{dC(t)}{dt} = -k_{\mathrm{off}} C(t) + k_{\mathrm{on}}(p_{\mathrm{TOT}} - C(t))X(t), \tag{6.4b}$$

in which the terms in the box represent the signal

$$s(t) = k_{\mathrm{off}} C(t) - k_{\mathrm{on}}(p_{\mathrm{TOT}} - C(t))X(t), \tag{6.4c}$$

that is, the retroactivity to the output. We can interpret s as a mass flow between the upstream and the downstream system. When $s = 0$, equation (6.4a) reduces to the dynamics of the isolated system given in equation (6.3).

The effect of the retroactivity s on the behavior of X can be very large (figure 6.9). This is undesirable in a number of situations where we would like an upstream system to "drive" a downstream one, as is the case, for example, when a biological oscillator has to time a number of downstream processes. If, due to the retroactivity, the output signal of the upstream process becomes too low or out of phase with the output signal of the isolated system (as in figure 6.9), the coordination between the oscillator and the downstream processes will be lost. Here we focus on the retroactivity to the output s.

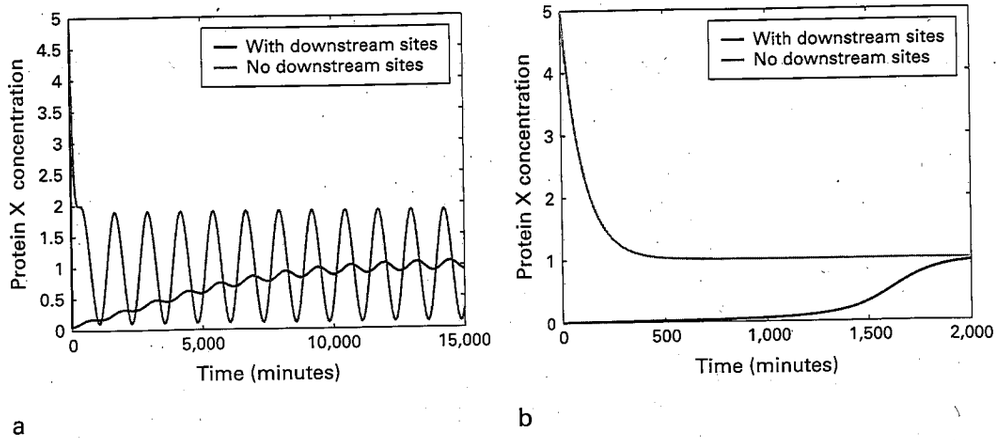

Figure 6.9
Dramatic effect of interconnection. Simulation results for the system in equation (6.4). The lighter line represents $X(t)$ as described by equations (6.3) (no retroactivity), while the darker line represents $X(t)$ obtained by equation (6.4) (with retroactivity). Both transient and longtime behaviors are different. Here $k(t) = 0.01(1 + \sin(\omega t))$ with $\omega = 0.005$ in panel a, and $\omega = 0$ in panel b, $k_{on} = 10$, $k_{off} = 10$, $\delta = 0.01$, $p_{TOT} = 100$ and $X(0) = 5$. The choice of protein decay rate (in min^{-1}) corresponds to a half-life of about one hour. The oscillations are chosen to have a period of about 12 times the protein half-life in accordance to what is experimentally observed in the synthetic clock of Atkinson et al. (2003).

In general, we will model retroactivity by a signal that travels from downstream to upstream. We thus model a system by adding an additional input, s, to model any change in the system dynamics that may occur upon interconnection with a downstream system. Similarly, we add to a system a signal r as another output to model the fact that, when such a system is connected downstream of another system, it will send upstream a signal that will alter the dynamics of the upstream system. More generally, we define a system S to have internal state x, two types of inputs (I), and two types of outputs (O): an input u (I), an output y (O), a *retroactivity to the input* r (O), and a *retroactivity to the output* s (I) (figure 6.10). We will thus represent a system S by the equations

$$\frac{dx(t)}{dt} = f(x(t), u(t), s(t)), \tag{6.5a}$$

$$y(t) = Y(x(t), u(t), s(t)), \tag{6.5b}$$

$$r(t) = R(x(t), u(t), s(t)), \tag{6.5c}$$

in which the functions f, Y, and R can take any form and the signals x, u, s, r, and y may be scalars or vectors. In such a formalism, the input-output model of the isolated system is recovered from equations (6.5) by eliminating r and setting $s = 0$.

Let system S_i have inputs u_i and s_i and outputs y_i and r_i. Let S_1 and S_2 be two systems with disjoint sets of internal states. We define the interconnection of an upstream system S_1 with a downstream system S_2 by setting $y_1 = u_2$ and $s_1 = r_2$. We will only consider the interconnections of systems that do not have internal states in common.

6.4.2 Quantification of the Retroactivity to the Output

An operative quantification of the retroactivity to the output can be obtained by exploiting the difference of time scales between the dynamics of the output of the upstream module and the dynamics of the input stage of the downstream module. This separation of time scales is always encountered in transcriptional circuits as discussed in section 6.3. We quantify the difference between the dynamics of X in the isolated system (6.3) and the dynamics of X in the connected system (6.4) by establishing conditions on the biological parameters under which the two systems exhibit similar behavior. This is achieved by exploiting the difference of time scales between the protein production and decay processes and the binding and unbinding process at the promoter p. By virtue of this separation of time scales, we can approximate system (6.4) by a one-dimensional system describing the evolution of X on the slow manifold (Kokotović et al., 1999). This reduced system takes the form

$$\frac{d\bar{X}(t)}{dt} = k(t) - \delta\bar{X}(t) + \bar{s}(t),$$

where $\bar{X}$ is an approximation of X and $\bar{s}$ is an approximation of s, which can be written as $\bar{s} = -\mathcal{R}(\bar{X})(k(t) - \delta\bar{X})$ with

$$\mathcal{R}(\bar{X}) = \frac{1}{1 + \dfrac{(1 + \bar{X}/k_d)^2}{p_{TOT}/k_d}}, \tag{6.6}$$

where k_d is the dissociation constant for transcription factor binding, $k_d = k_{off}/k_{on}$ (see Del Vecchio et al., 2008b, for details). The expression $\mathcal{R}(\bar{X})$ quantifies the retroactivity to the output on the dynamics of X after a fast transient, when we approximate X with $\bar{X}$ in the limit where $\delta/k_{off} \approx 0$. The retroactivity measure is thus low if the affinity of the binding sites p is small (k_d large) or if the signal $X(t)$ is large enough compared to p_{TOT}. Thus the form of $\mathcal{R}(\bar{X})$ provides an operative quantification of the retroactivity: such an expression can in fact be evaluated once the association and dissociation constants of X to p, the concentration of the binding sites p_{TOT}, and the range of operation of the signal $\bar{X}(t)$ that travels across the interconnection are all known. Thus the modularity assumption introduced in section 6.2 holds if the value of $\mathcal{R}(\bar{X})$ is low enough.

Figure 6.10
System S input and output signals. The dotted arrows denote signals originating by retroactivity upon interconnection.

6.5 Insulation Devices to Enforce Modularity

Of course, it is not always possible to design an interconnection such that the retroactivity is low. This is, for example, the case of an oscillator that has to time a downstream load: in general, the load cannot be included in the design as the oscillator must perform well in the face of unknown and possibly variable loads. That said, as with electrical circuits, one can design a device, to be placed between the oscillator and the load, such that the output of the device is unaffected by the load and the device itself does not affect the behavior of the upstream oscillator. Specifically, consider a system S as shown in figure 6.10 that takes u as input and gives y as output, designed in such a way that

1. the retroactivity r to the input is very small;
2. the effect of the retroactivity s to the output on the internal dynamics of the system is very small;
3. its input-output relationship is about linear.

A system like this is said to enjoy the *insulation* property and will be called an insulation device. Indeed, it will not affect an upstream system because $r \approx 0$ and it will keep the same output signal y *independently* of any connected downstream system. Other researchers have considered the insulation from external perturbations and robustness properties of amplifiers in the context of biochemical networks (Sauro and Ingalls, 2007; Sauro and Kholodenko, 2004). Here we revisit the amplifier mechanism in the context of gene transcriptional networks with the objective of mathematically and computationally demonstrating how suitable biochemical realizations of such a mechanism can attain properties 1, 2, and 3.

In electrical circuits, the standard insulation device is the operational amplifier, or OPAMP. In electronic amplifiers, r is very small because the input stage of an OPAMP absorbs almost no current. This way, there is no voltage drop across the output impedance of an upstream voltage source. Equation (6.6) quantifies the effect of retroactivity on the dynamics of X as a function of biochemical parameters that characterize the interconnection mechanism with a downstream system. These parameters are the affinity of the binding site $1/k_d$, the total concentration of the

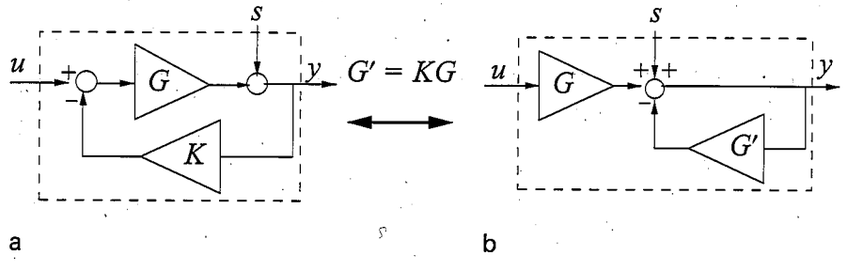

Figure 6.11
(a) Basic feedback/amplification mechanism by which amplifiers attenuate the effect of the retroactivity to the output s. (b) Alternative representation of the same mechanism, which will be employed to design biological insulation devices.

promoter, p_{TOT}, and the level of the signal $X(t)$. To reduce retroactivity, we can choose k_d large (low affinity) and p_{TOT} small, for example. Having a small value for p_{TOT}, low affinity, or both implies that there is a small "flow" of protein X toward its target sites. Thus we can say that a low retroactivity to the input is obtained when the "input flow" to the system is small. This interpretation establishes a nice analogy to the electrical case, in which low retroactivity to the input is obtained by a low input current.

In an electronic amplifier, the effect of the retroactivity to the output s on the amplifier behavior is reduced to almost zero by virtue of a large (theoretically infinite) input amplification gain and a negative output feedback. Such a mechanism can be illustrated in its simplest form by figure 6.11a, which is very well known to control engineers. For simplicity, we have assumed in such a diagram that the retroactivity s is just an additive disturbance. That the effect of the retroactivity s to the output is negligible for large gains G can be verified through the following simple computation. The output y is given by $y = G(u - Ky) + s$, which leads to

$$y = u\frac{G}{1+KG} + \frac{s}{1+KG}.$$

As G grows, y tends to u/K, which is independent of the retroactivity s. To attenuate the retroactivity effect at the output of a component one (1) amplifies the input of the component through a large gain and (2) applies a large negative output feedback (figure 6.11b). We next illustrate this idea in the context of the transcriptional example. Consider the approximated dynamics of X. Let us assume that we can apply a gain G to the input $k(t)$ and a negative feedback gain G' to X with $G' = KG$. This leads to the new differential equation for the connected system given by

$$\frac{dX(t)}{dt} = (Gk(t) - (G' + \delta)X(t))(1 - \mathcal{R}(X(t))). \tag{6.7}$$

It can be shown (see Del Vecchio et al., 2008a, for details) that, as G and thus as G' grows, the signal $X(t)$ generated by the connected system given by equation (6.7) becomes close to the solution $\bar{X}(t)$ of the isolated system

$$\frac{d\bar{X}(t)}{dt} = Gk(t) - (G' + \delta)\bar{X}(t); \tag{6.8}$$

that is, the presence of the disturbance term $\mathcal{R}(X)$ will not significantly affect the time behavior of $X(t)$. A key question arises: How can we obtain a large amplification gain G and a large negative feedback G' in a biological insulation component? This question is addressed in the following section, in which we show that a simple phosphorylation/dephosphorylation cycle has remarkable insulation properties (for additional designs of biomolecular insulation devices, see Del Vecchio et al., 2008b).

6.5.1 A Biomolecular Realization of an Insulation Device through Protein Phosphorylation

In this design, we propose to obtain input amplification through a fast phosphorylation reaction and negative feedback through a fast dephosphorylation reaction. In particular, this is realized by having the input Z activate the phosphorylation of a protein X, which is available in the system in abundance. That is, Z is a kinase for a protein X. The phosphorylated form of X, X_p, binds to the downstream sites, whereas X does not. A negative feedback on X_p is obtained by having a phosphatase Y activate the dephosphorylation of protein X_p. Protein Y is also available in abundance in the system. This mechanism is depicted in figure 6.12. A similar design has been proposed by Sauro and Kholodenko (2004) and Sauro and Ingalls (2007), in which a MAPK cascade plus a negative feedback loop that spans the length of the MAPK cascade is considered as a feedback amplifier. Our design is much simpler, involving only one phosphorylation cycle and requiring no additional feedback loop.

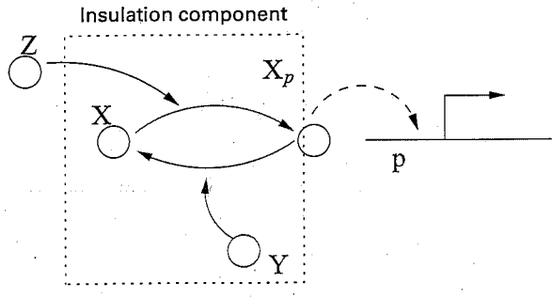

Figure 6.12
Insulation device (*dotted box*).

To convey the idea of how this device realizes the insulation function, we consider the one-step reaction model for the phosphorylation reactions analyzed by Heinrich et al. (2002):

$$Z + X \xrightarrow{k_1} Z + X_p \quad \text{and} \quad Y + X_p \xrightarrow{k_2} Y + X.$$

We assume that there is plenty of protein X and of phosphatase Y in the system and that these quantities are conserved. The conservation of X gives $X + X_p + C = X_{TOT}$, in which X is the inactive protein, X_p is the phosphorylated protein that binds to the downstream sites p, and C is the complex of the phosphorylated protein X_p bound to the promoter p. The X_p dynamics can be described by the first equation in the following model:

$$\frac{dX_p(t)}{dt} = k_1 X_{TOT} Z(t)\left(1 - \frac{X_p(t)}{X_{TOT}} - \boxed{\frac{C(t)}{X_{TOT}}}\right) - k_2 Y(t) X_p(t) + \boxed{k_{off} C(t) - k_{on} X_p(t)(p_{TOT} - C(t))}, \tag{6.9}$$

$$\frac{dC(t)}{dt} = -k_{off} C(t) + k_{on} X_p(t)(p_{TOT} - C(t)). \tag{6.10}$$

The boxed terms represent the retroactivity s to the output of the insulation system of figure 6.12. For a weakly activated pathway, $X_p \ll X_{TOT}$ (Heinrich et al., 2002). Also, if we assume that the total concentration of X is large compared to the concentration of the downstream binding sites, that is, $X_{TOT} \gg p_{TOT}$, equation (6.9) is approximated by

$$\frac{dX_p(t)}{dt} = k_1 X_{TOT} Z(t) - k_2 Y(t) X_p(t) + k_{off} C(t) - k_{on} X_p(t)(p_{TOT} - C(t)).$$

If we denote $G = k_1 X_{TOT}$ and $G' = k_2 Y$ and again exploit the difference of time scales between the X_p dynamics and the C dynamics after a fast initial transient, we can approximate the dynamics of X_p accurately by

$$\frac{dX_p(t)}{dt} = (G(t)Z(t) - G'(t)X_p(t))(1 - \mathcal{R}(X_p(t))), \tag{6.11}$$

in which $\mathcal{R}(X_p)$ is the measure of the retroactivity s to the output after a short transient. Therefore, for G and G' large enough, $X_p(t)$ is well described by the isolated system $dX_p(t)/dt = G(t)Z(t) - G'(t)X_p(t)$. As a consequence, the effect of the retroactivity to the output s is attenuated by increasing $k_1 X_{TOT}$ and $k_2 Y$. That is, to obtain large input and feedback gains, one should have large phosphorylation/dephosphorylation rates, large amounts of protein X and phosphatase Y in the system, or both.

To highlight the roles of the various parameters for attaining the insulation properties, we can consider a more complex model for the phosphorylation and dephosphorylation reactions and perform parametric analysis. In particular, let us consider a two-step reaction model such as that of Huang and Ferrell (1996). According to this model, we have the following two reactions for phosphorylation and dephosphorylation, respectively:

$$X + Z \underset{\beta_2}{\overset{\beta_1}{\rightleftharpoons}} C_1 \xrightarrow{k_1} X_p + Z, \tag{6.12}$$

and

$$Y + X_p \underset{\alpha_2}{\overset{\alpha_1}{\rightleftharpoons}} C_2 \xrightarrow{k_2} X + Y, \tag{6.13}$$

in which C_1 is the complex of protein X with kinase Z and C_2 is the complex of phosphatase Y and protein X_p. Additionally, we have the conservations $Y_{TOT} = Y + C_2$, $X_{TOT} = X + X_p + C_1 + C_2 + C$, because proteins X and Y are not degraded. The differential equations modeling the insulation system of figure 6.12 thus become

$$\frac{dZ(t)}{dt} = k(t) - \delta Z(t)$$

$$\boxed{-\beta_1 Z(t) X_{TOT}\left(1 - \frac{X_p(t)}{X_{TOT}} - \frac{C_1(t)}{X_{TOT}} - \frac{C_2(t)}{X_{TOT}} - \boxed{\frac{C(t)}{X_{TOT}}}\right) + (\beta_2 + k_1) C_1(t)}, \tag{6.14}$$

$$\frac{dC_1(t)}{dt} = -(\beta_2 + k_1) C_1(t)$$

$$+ \beta_1 Z(t) X_{TOT}\left(1 - \frac{X_p(t)}{X_{TOT}} - \frac{C_1(t)}{X_{TOT}} - \frac{C_2(t)}{X_{TOT}} - \boxed{\frac{C(t)}{X_{TOT}}}\right), \tag{6.15}$$

$$\frac{dC_2(t)}{dt} = -(k_2 + \alpha_2) C_2(t) + \alpha_1 Y_{TOT} X_p(t)\left(1 - \frac{C_2(t)}{Y_{TOT}}\right), \tag{6.16}$$

$$\frac{dX_p(t)}{dt} = k_1 C_1(t) + \alpha_2 C_2(t) - \alpha_1 Y_{TOT} X_p(t)\left(1 - \frac{C_2(t)}{Y_{TOT}}\right)$$

$$+ \boxed{k_{off} C(t) - k_{on} X_p(t)(p_{TOT} - C(t))}, \tag{6.17}$$

$$\frac{dC(t)}{dt} = -k_{off} C(t) + k_{on} X_p(t)(p_{TOT} - C(t)), \tag{6.18}$$

in which C is the complex of X_p with the downstream promotor p, and the expression of gene z is controlled by a promoter with activity $k(t)$. The terms in the large box in equation (6.14) represent the retroactivity r to the input, while the terms in the small box in equation (6.14) and in the boxes of equations (6.15) and (6.17) represent the retroactivity s to the output. A detailed analysis of the system in equations (6.14–6.18) also provides analytical relationships among the parameters that indicate how to obtain small retroactivity to the input r and linear input-output behavior (see Del Vecchio et al., 2008b, for details). As for the simplified model, we have again that $G \propto k_1 X_{TOT}$ and $G' \propto k_2 Y_{TOT}$.

The system in equations (6.14–6.18) was simulated with and without the downstream binding sites p, that is, with and without the terms in the small box of equation (6.14) and in the boxes of equations (6.17) and (6.15). This analysis highlights the effect of the retroactivity to the output s on the dynamics of X_p. The simulations

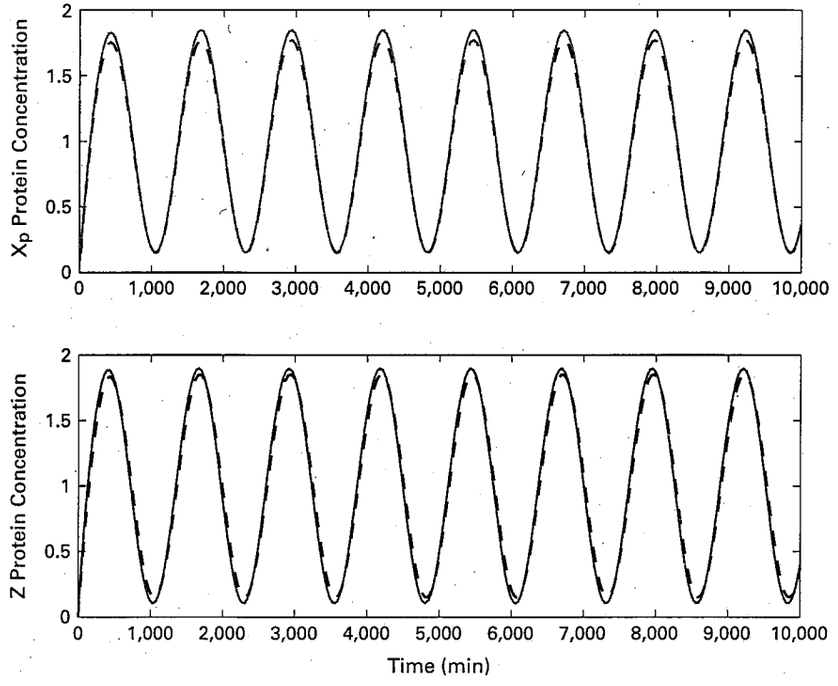

Figure 6.13
Simulation results for system described by equations (6.14–6.18). In all plots, $p_{TOT} = 100$, $k_{off} = k_{on} = 10$, $\delta = 0.01$, $k(t) = 0.01(1 + \sin(\omega t))$, and $\omega = 0.005$. In subplots A and B, $k_1 = k_2 = 50$, $\alpha_1 = \beta_1 = 0.01$, $\beta_2 = \alpha_2 = 10$, and $X_{TOT} = Y_{TOT} = 1{,}500$. Panel a shows the signal $X_p(t)$ without the downstream binding sites p (*solid line*) and the same signal with the downstream binding sites p (*dashed line*). The small error shows that the effect of the retroactivity to the output s is attenuated very well. Panel b shows the signal $Z(t)$ without X, to which Z binds (*solid line*), and the same signal $Z(t)$ with X present in the system (*dashed line*). The small error confirms a small retroactivity to the input.

validate our theoretical study indicating that, when $X_{TOT} \gg p_{TOT}$ and the time scales of phosphorylation/dephosphorylation are much faster than the time scale of decay and production of the protein Z, the retroactivity to the output s is very well attenuated (figure 6.13a). Similarly, the time behavior of Z was simulated with and without the terms in the large box of equation (6.14), that is, with and without X to which Z binds, to verify whether the insulation component exhibits retroactivity to the input r. In particular, the accordance of the behaviors of $Z(t)$ with and without its downstream binding sites on X (figure 6.13b) indicates that there is no substantial retroactivity to the input r generated by the insulation device.

6.6 Conclusions and Future Challenges

We have reviewed some design methods employed in synthetic biology that rely on the modularity assumption according to which modules maintain their dynamic behavior unchanged upon interconnection. By virtue of this assumption, one can first characterize a module in isolation and then predict the behavior of a circuit directly by the behavior of its composing modules. This is a powerful approach to circuit design, which is employed also in other engineering areas, such as electronics. We pointed out, however, that, just as in several other engineering systems, because of loading effects at interconnections, modularity does not necessarily hold in biomolecular systems. As with historical developments in electronics, where researchers focused on characterizing impedance (loading) effects and on counteracting them with the aid of operational amplifiers, similar efforts are currently taking place in the area of biomolecular circuit design.

Biomolecular circuit design presents many challenges to systems and control engineers. We need to address problems of loading effects at interconnections between general biomolecular systems, such as signaling systems, as opposed to purely transcriptional systems. These include not only the effects of loads but also impedance-matching problems and frequency sensitivity analysis. We need to consider energetic constraints in the design of active devices such as insulation devices, which could possibly impose tough requirements on the cell, as well as possible trade-offs between the need for large gains in the design of insulation devices and the effects of such large gains on biological noise. Finally, to these engineering challenges must be added the scientific challenge of uncovering design principles that are already employed by natural systems for coping with similar problems.

Note

1. A significant portion of the contents of this chapter appeared in Del Vecchio et al., 2008b.