

Mads Kærn and Ron Weiss

In parallel with the development of high-throughput technologies fueling systems biology, advances in modeling of biological systems and in synthesis of long DNA fragments with arbitrary nucleotide sequences have fostered the emergence of a nascent field termed synthetic biology. At its core, this field uses recombinant DNA manipulation techniques to design and embed complex “programmed” functions into living organisms. An important notion that pervades most of the work in synthetic biology is the use of mathematical models for forward design. As such, systems and synthetic biology can be viewed as being two sides of the same coin. While systems biology attempts to unravel how the set of instructions encoded by an organism’s DNA orchestrates its phenotypical complexity, synthetic biology aims to create cells with desirable behaviors through the integration of additional instructions. This can be achieved by first investigating which network architectures support the desired outcome and then augmenting the genotype accordingly. The construction of synthetic gene regulatory systems can thus help understand natural systems by complementing approaches in which quantitative analysis is used to elucidate “design principles” underlying the functioning of natural intracellular networks. Moreover, synthetic systems provide excellent examples of the direct link between theoretical modeling and biological reality.

13.1 Introduction

During the last few decades, the ability to isolate, sequence, and manipulate DNA has led to tremendous advances in genetic engineering with numerous benefits to science, agriculture, and medicine. Typically, genetic engineering is used to endow a genetically modified organism with a novel trait, such as resistance to certain pesticides or the ability to efficiently synthesize pharmacological molecules, for example by transferring a gene from another organism. Gene therapy is another example. There, a trait lost due to a nonfunctional endogenous gene is typically

recovered by inserting a normal copy of the gene at a non-specific location within the genome. Synthetic biology can be viewed as a natural extension of such single-gene approaches in the sense that entire systems are inserted into the genome of the host cell.

So far, efforts in synthetic biology have included the construction of novel gene regulatory networks, signal transduction pathways, metabolic pathways, synthetic multicellular systems, engineered sensory proteins, and the regulation of proteins that control intrinsic cell functions. An excellent introductory review of the many different aspects of synthetic biology is given by Benner and Sismour (2005). Here, we focus on the mathematical design and experimental implementation of selected synthetic gene regulatory networks that embody important architectural properties. Prerequisites for designing and implementing synthetic gene regulatory networks include understanding how transcriptional regulation works, how transcription factor proteins regulate the expression of each other within networks, and knowledge of recombinant DNA technologies. An excellent introduction to the latter is given by Nicholl (1994).

General aspects of transcriptional regulation and how transcription is modeled are discussed in sections 13.2 and 13.3, respectively. The remaining sections highlight how synthetic gene regulatory systems have been designed and implemented in the bacterium *Escherichia coli* based on network models constructed from phenomenological mathematical descriptions of transcriptional regulation. In sections 13.4 and 13.5, we discuss linear transcriptional networks and feedforward networks, respectively. In section 13.6, we provide examples of networks that support bistability and oscillations by incorporating feedback control. These systems demonstrate how some of the principles investigated in chapter 6 have been used to create living cells with complex dynamical properties.

13.2 Transcriptional Regulatory Modules

In order to engineer gene regulatory systems, it is necessary to appreciate some of the basic elements of gene regulation. Natural genetic circuits are typically described as circuits of interconnected modules consisting of interacting proteins, DNA, RNA, and small molecules that regulate the transcription of genes into mRNA, the translation of mRNA into polypeptides, and the biological activity of the expressed proteins. While the abundance of an expressed protein can be controlled by many different mechanisms, the regulation of gene transcription is one of the most common. In prokaryotes, this type of control is often mediated through transcription factor proteins that alter the ability of the RNA polymerase to bind to and initiate transcription from promoter regions located upstream of the regulated genes.

Prokaryotic transcriptional regulatory modules often consists of four elements: a promoter region, the gene (or genes) expressed from that promoter, the transcription factor proteins that regulate the expression level, and additional regulatory

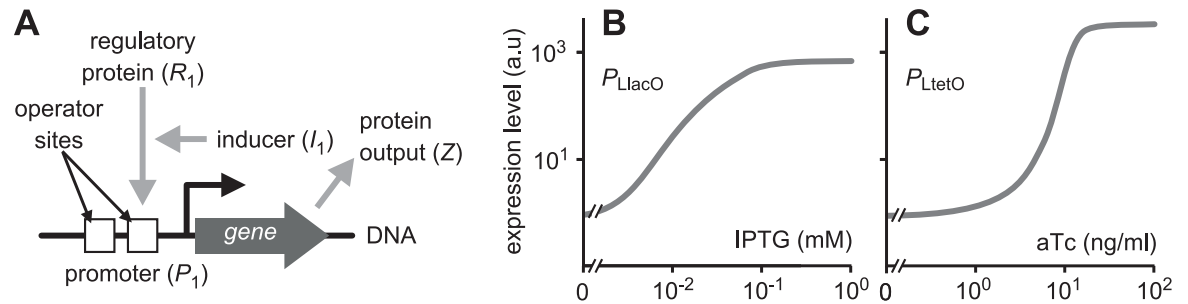


Figure 13.1 **A.** Architecture of a prototypical transcriptional regulatory module. **B–C.** Population-averaged signal-response curves measured in *E. coli* cells expressing a reporter gene from LacI/ P_{LlacO} and TetR/ P_{LtetO} modules with fixed concentrations of LacI or TetR and varying concentrations of the inducers isopropyl- β -D-thiogalactoside (IPTG) and anhydrotetracycline (aTc), respectively. Based on experimental data from (Lutz and Bujard, 1997).

molecules that modulate the activity of the transcription factors. A schematic layout of such modules is shown in figure 13.1A. The expression from the regulated promoter can be measured in single cells by expressing a reporter gene, such as the *gfp*, *yfp*, or *cfp* genes encoding green (GFP), yellow (YFP) and cyan (CFP) fluorescent protein, respectively (see chapter 10).

Transcriptional regulatory proteins increase (activators) or decrease (repressors) the probability that a gene is transcribed into mRNA by binding to stretches of DNA within or near promoter regions referred to as operators or *cis*-regulatory elements (figure 13.1A). While activators may facilitate the binding of RNA polymerase to the promoter, repressors often exert their function by competing with the RNA polymerase for promoter access. Transcription from a promoter containing appropriate *cis*-regulatory elements can thus be controlled by up- or down-regulating the cellular abundance of the corresponding transcription factor proteins. In some cases, external control over such *in vivo* signals is provided by small molecules called inducers. These molecules typically function by modulating the activity of a transcription factor protein. Specifically, when the inducer binds to the protein, it causes an alteration in its three-dimensional structure that increases or decreases the affinity between the protein and its cognate *cis*-regulatory elements. Varying the inducer concentration thus provides a means of regulating transcription without altering cellular protein abundances directly.

Figures 13.1B and 13.1C illustrate how expression of a reporter protein from two engineered transcriptional regulatory modules, LacI/ P_{LlacO} and TetR/ P_{LtetO} , is modulated by the inducers isopropyl- β -D-thiogalactoside (IPTG) and anhydrotetracycline (aTc), respectively. The P_{LlacO} and P_{LtetO} promoters are obtained by inserting *lacO* and *tetO* operator sequences, corresponding to the binding sites of LacI and TetR, respectively, into the P_L promoter normally repressed by the protein CI. In both cases, the signal-response curve, in other words, the relationship between the regulatory input signal (the inducer concentration) and the output signal (the abundance of the reporter protein), is highly nonlinear and sigmoidal. The endogenous *E. coli* promoters P_{lac} and P_{tet} , which are repressed by LacI and

Table 13.1 Transcription regulatory modules used frequently to construct synthetic gene regulatory networks in *E. coli*. TetR and LacI are repressors that are inactivated by their inducers. LuxR is an AHL-dependent activator. CI is generally a repressor, but activates transcription from the P_{RM} promoter.

Regulatory protein	Regulated promoters	Inducer
TetR	P_{tet} , P_{LtetO}	tetracycline
LacI	P_{lac} , P_{LlacO} , P_{trc}	lactose, IPTG
CI	P_L , P_R , P_{RM} , P_{luxOR}	
LuxR	P_{lux} , P_{luxOR}	acyl-homoserine lactones (AHL)

TetR, respectively, respond to induction in a similar fashion. Additional transcriptional regulatory modules used frequently in synthetic biology are summarized in table 13.1.

13.3 Modeling Transcriptional Modules

In the remaining sections of this chapter, we discuss how simulation and analysis of mathematical models have been employed to forward engineer *E. coli* cells with novel characteristics and sophisticated computational capabilities by interconnecting the modules in table 13.1 into larger networks. We use a convenient abstraction to model these biochemical networks with ordinary differential equations (ODEs) that include basal expression of a protein, protein decay, and Hill function descriptions of gene regulation (see chapter 6).

In general form, the ODE that models the output Z of a genetic module given the regulatory input S is given by:

$$\frac{d[Z]}{dt} = k' + \frac{k \cdot (S^n/K^n)^\mu}{1 + (S^n/K^n)} - d \cdot [Z] \quad (13.1)$$

where the parameter μ is used to distinguish between the cases of repression ($\mu = 0$) and activation ($\mu = 1$) of transcription by S (see for example, Kuznetsov et al. (2005) for details). The constants K and n are the Hill constant and Hill coefficient, respectively. The Hill constant gives the value of the input signal that yields 50% response, and the Hill coefficient gives the slope of the signal-response curve at this input signal. The parameter d is the rate constant associated with the decay of the output reporter protein. Additionally, the parameters k and k' are the rate constants associated with signal-independent (basal) and signal-dependent gene expression. The values of k and k' are typically correlated, and this interdependence is often modeled by setting $k' = a \cdot k$ with $0 \leq a < 1$. With this relationship, the steady state solution of equation 13.1 is given by:

$$[Z]_{ss} = \frac{k}{d} \left(a + \frac{(S^n/K^n)^\mu}{1 + (S^n/K^n)} \right) \quad (13.2)$$

Hence, in steady state, the cellular abundance of the reporter protein reflects the relationship between the regulatory signal and the transcription rate modeled by the Hill function.

It is noted that equation 13.1 models transcription and translation as a single step. Because the separation of transcription and translation introduces response delays, it can be important in models of temporal dynamics to include mRNA as an independent variable. In this case, a single-input transcriptional regulatory module is described by the ODEs:

$$\begin{aligned}\frac{d[M]}{dt} &= a \cdot k_{tr} + \frac{k_{tr} \cdot (S^n/K^n)^\mu}{1 + (S^n/K^n)} - d_M \cdot [Z] \\ \frac{d[Z]}{dt} &= k_{tl}[M] - d \cdot [Z]\end{aligned}\tag{13.3}$$

where $[M]$ is the mRNA concentration, k_{tr} is the rate constant associated with transcription, k_{tl} is the rate constant associated with translation and d_M is the mRNA decay constant. Equation 13.1 is obtained from equation 13.3 by invoking a steady state assumption for $[M]$ and defining the constant k by $k = k_{tr}k_{tl}/d_M$. Hence, modeling transcription and translation as a single step does not change the steady state solution in equation 13.2.

Equation 13.1, or equation 13.3 when mRNA is included, are used to model both the effect of changing the intracellular concentration of a regulatory protein and the extracellular concentration of an inducer. For example, the effect of varying the concentration of a repressor R is modeled by setting the input signal equal to the repressor concentration, $S = [R]$, with $\mu = 0$. When the concentration of the repressor is constant, the effect of varying the concentration of its inducer I is modeled by setting $S = [I]$ and $\mu = 1$. For example, the steady state signal-response curves in Figs. 13.1B and 13.1C for induction of the LacI/ P_{LacO} and the TetR/ P_{LtetO} modules can be modeled using equation 13.2 with $\mu = 1$ and the concentrations of IPTG and aTc defining the signal S , respectively. Other input-output functions are also possible depending on the regulatory role of the protein and how the inducer affects the activity of this protein. In cases where both repressor and inducer concentrations vary, the signal S is the concentration of active repressor molecules. This signal is modeled by setting $S = [R_T]/(K_I^{n_I} + [I]^{n_I})$ with $[R_T]$ being the total repressor concentration, and K_I and n_I the Hill constant and coefficient associated with the repressor-inducer interaction, respectively.

For the purpose of network modeling, we will use the following notations: Each transcriptional regulator protein is given an index $i = 1, 2, \dots, N$. The concentration of the transcription factor protein is given by $[R_i]$, and its inducer, if present, by $[I_i]$. The gene and mRNA that encode the regulatory protein R_i are denoted $r(i)$ and M_i , respectively. The rate constant associated with the decay of protein R_i is given by d_i . Promoters are identified as follows: P denotes a constitutively active promoter, P_i a promoter regulated by the protein R_i , and P_{ij} a promoter regulated by the proteins R_i and R_j . The parameters characterizing the transcription from each promoter are identified by the same index as the promoter for the parameters

k and a (or k'), and by the protein index for the Hill constant K and the Hill coefficient n .

13.4 Linear Networks

Linear transcriptional regulatory networks consist of modules placed in series with the output of one module acting as the input to the next module. In the simplest case, a linear network is composed of two modules and one regulatory step. The LacI/ P_{LacO} and TetR/ P_{TetO} modules discussed in section 13.2 are examples of such one-step transcriptional cascades because the transcriptional regulator (R_1 in figure 13.1) is expressed at high constant levels from a constitutively active promoter. For clarity, the constitutive promoter (that is, the first transcription module) is omitted from the diagram in figure 13.1. The construction and analysis of longer transcriptional cascades, which will be discussed next, is useful for determining how information flows through transcriptional networks and can help better understand the rules of module composition. For example, cascades comprised of two and three regulatory steps have been engineered with the purpose of investigating time delays, ultrasensitivity in signal-response relationships and stochasticity in transcriptional regulation (see, for example, Blake et al. (2003); Hooshangi et al. (2005); Rosenfeld et al. (2005); Pedraza and van Oudenaarden (2005)).

13.4.1 Two-Step Cascades

Figure 13.2 depicts the schematics of a two-step linear repressor cascade obtained by adding a third transcriptional module to the one-step cascade. The first module comprises the promoter P with no regulatory inputs; the second module the repressor R_1 , its inducer I_1 , and the promoter P_1 . The third module comprises the R_2 repressor and the P_2 promoter. This configuration provides a mechanism to measure the behavior of the R_2/P_2 module. The constitutive promoter P drives expression of the R_1 repressor, which in turn, inhibits the expression of R_2 . The inducer I_1 can thus be used to determine the input signal to the R_2/P_2 regulatory module by modulating the cellular abundance of repressor R_2 .

Using equation 13.1 with $\mu = 1$ and $S = [I_1]$ to model the R_1/I_1 -dependent expression of R_2 and with $\mu = 0$ and $S = [R_2]$ to model the inhibition by R_2 of expression from P_2 , the ODEs describing the two-step linear network are given by:

$$\begin{aligned} \frac{d[R_2]}{dt} &= a_1 \cdot k_1 + \frac{k_1 \cdot ([I_1]/K_1)^{n_1}}{1 + ([I_1]/K_1)^{n_1}} - d_2 \cdot [R_2] \\ \frac{d[Z]}{dt} &= a_2 \cdot k_2 + \frac{k_2}{1 + ([R_2]/K_2)^{n_2}} - d \cdot [Z] \end{aligned} \quad (13.4)$$

where the meaning of the parameters were defined in section 13.3. Notice that it is not necessary to include an equation for R_1 because its steady state level is constant. As discussed in section 13.3, the combined regulatory activity of R_1 and I_1 can be

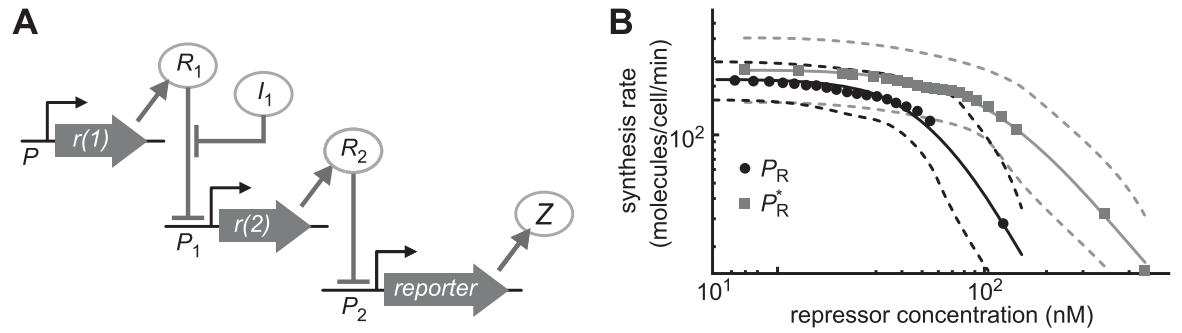


Figure 13.2 **A.** Architecture of a two-step repressor cascade. **B.** Population-averaged rates of reporter protein synthesis from the P_R promoter (black points) and the P_R^* promoter (grey points) measured at the single-cell level using a two-step repressor cascade. The broken curves give the standard deviation associated with the measured synthesis rate, and full curves the fit to a Hill function with $\mu = 0$. The fitted parameter values are: $n = 2.4 \pm 0.3$, $K = 55 \pm 10$ nM, $k = 220 \pm 15$ min^{-1} for the P_R promoter, and $n = 1.7 \pm 0.3$, $K = 120 \pm 25$ nM, $k = 255 \pm 40$ min^{-1} for the P_R^* promoter (Rosenfeld et al., 2005).

captured phenomenologically in one Hill function to model the relationship between the inducer concentration and the expression from the regulated promoter.

The steady state solution of equation 13.4 is given by:

$$\begin{aligned} [R_2]_{ss} &= \frac{k_1}{d_2} \left(a_1 + \frac{([I_1]/K_1)^{n_1}}{1 + ([I_1]/K_1)^{n_1}} \right) \\ [Z]_{ss} &= \frac{k_2}{d} \left(a_2 + \frac{1}{1 + ([R_2]_{ss}/K_2)^{n_2}} \right) \end{aligned} \quad (13.5)$$

In terms of the overall response of this network, the steady state solution predicts that the presence of inducer (high input) results in repression of P_2 (high R_2 , low output) and the absence of inducer (low input) allows transcription from the P_2 promoter (low R_2 , high output).

13.4.2 Characterizing Module Input-Output Functions

The network illustrated in figure 13.2 and described by the model in equation 13.4 can be implemented using different repressor/promoter pairs. Elowitz and colleagues (Rosenfeld et al., 2005) implemented a version using the aTc-inducible TetR/ P_{tet} module to characterize a CI/ P_R repressor module driving CFP ($R_1 = \text{TetR}$, $I_1 = \text{aTc}$, and $R_2 = \text{CI}$, $Z = \text{CFP}$). In this implementation, the *cI* gene is fused with the *yfp* gene to synthesize a yellow-fluorescent variant of the CI protein. This dual-color labeling allows for simultaneous measurements of the input and output signals in single cells. Additionally, using time-lapse microscopy to determine the rate of change in fluorescence, the dependency of the rate of protein synthesis on the repressor concentration can be determined at the level of single cells. The system thus enables a direct investigation of the suitability of the Hill function in equation 13.1 as a model of the most fundamental signal-response relationship in gene regulatory systems.

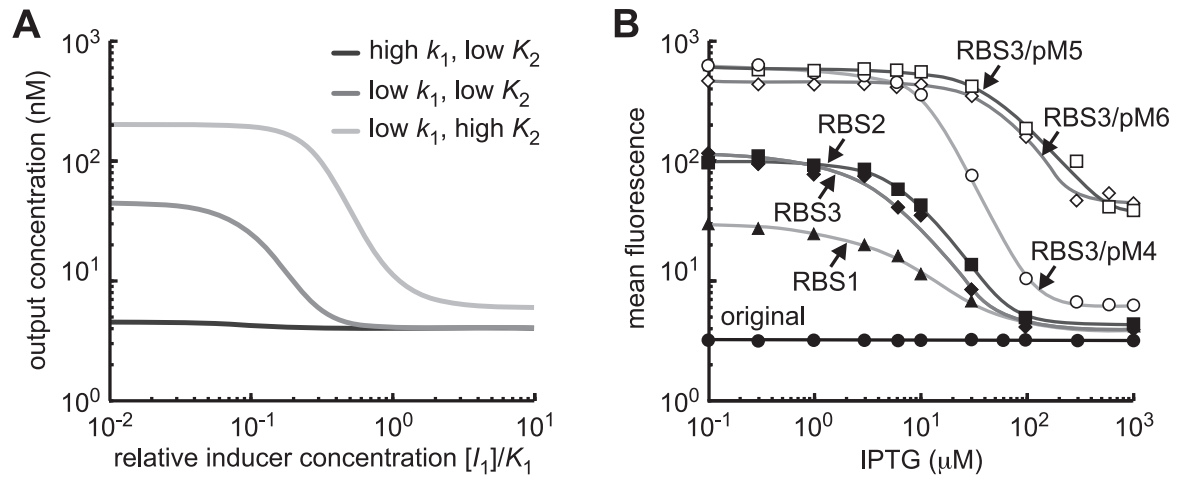


Figure 13.3 **A.** Predicted effects on the signal-response curve of the two-step repressor cascade of decreasing the rate of repressor R_2 synthesis (from $k_1 = 1000$ nM/min to $k_1 = 100$ nM/min) and increasing the Hill constant (from $K_2 = 10$ nM to $K_2 = 100$ nM). Other parameter values are $a_1 = a_2 = 0.02$, $d_1 = d_2 = 0.1$ nM/min, $k_2 = 20$ nM/min, and $n = 2$. **B.** Experimentally measured signal-response in two-step repressor cascades containing altered ribosome binding sites (RBS) of repressor-encoding mRNA to change k_1 , or mutations in the regulated promoter (pM) to change K_2 . Based on experimental data from (Weiss and Basu, 2002).

Figure 13.3B illustrates the experimentally observed relationship between the concentration of the CI-YFP protein and the population-averaged rate of CFP synthesis for the P_R promoter and a variant of this promoter, designated P_R^* , where one of the CI binding-sites is mutated. Also included are the standard deviations associated with the average protein synthesis rates and the signal-response curves obtained by fitting the data to Hill functions.

13.4.3 Matching Kinetic Characteristics

While the two-step cascade composed of the TetR/ P_{tet} and CI/ P_R modules exhibits a useful inverse sigmoidal signal-response relationship, it is often the case that coupling transcriptional regulatory modules does not yield the desired behavior. Another version of the same network uses the LacI/ P_{lac} pair as the inducible module to control the input to the CI/ P_R module. However, when initially assembled, no fluorescence was observed from cells harboring the network regardless of whether the inducer, in this case IPTG, is absent or present. Apparently, even with maximum repression of the P_{lac} promoter, CI is synthesized at a sufficiently high level to fully repress transcription from the P_R promoter. Unfortunately, our models are presently not sufficiently accurate to predict such mismatch problems partially because accurate *in vivo* parameter values are difficult to obtain. Hence, it is often necessary to first construct a network, and then use modeling tools to guide the correction and fine-tuning of its behavior.

In order to overcome impedance mismatch problems, one can mutate genetic elements until the desirable network response is obtained. Starting with a non-

functional or non-optimal network, such mutations can be introduced to affect biological parameters identified by model analysis as most likely to yield the desired behavior. For example, Feng et al. (2004) showed how to use global sensitivity analysis to determine the best genetic targets for mutations that could make the two-step $\text{LacI}/P_{\text{lac}}$, CI/P_{R} cascade functional. The steady state model in equation 13.5 predicts, as shown in figure 13.3A, that decreasing the value of the maximal repressor synthesis rate k_1 or increasing the Hill constant K_2 should confer a non-responsive network with the desired network properties. The Hill constant can be modified by mutating one of the CI-binding sites within the P_{R} promoter to lower the CI-binding affinity, and the maximal CI synthesis rate k_1 can be changed by mutating the ribosome-binding site (RBS) on the CI-encoding mRNA.

That the model correctly predicts the genetic mutations required to obtain a functional network is shown in figure 13.3B. The experimental results are obtained with three different *cI*-RBS sequences yielding lower translation efficiencies than the original RBS (Weiss and Basu, 2002). The plots show that the systems with the weakened RBS are able to respond to induction with IPTG, in agreement with the model predictions. Also shown are the effects of introducing mutations into the CI-binding site within the P_{R} promoter. These mutations are combined with the weakest RBS in order to optimize the response.

13.4.4 Interfacing Transcriptional Modules

Once the kinetic characteristics of the individual transcriptional regulatory modules are appropriately matched, they can be coupled together into larger networks. This can be accomplished by combining modules at random (Guet et al., 2002) or rationally to achieve a specific network property. Perhaps the simplest extension of the two-step cascade is to add an additional repressor module to form a linear three-step network (figure 13.4A). The experimental investigation of this cascade highlights interesting properties that are important for the understanding of the more complex systems discussed in the sections to follow.

An implementation of the three-step linear repressor cascade uses the $\text{TetR}/P_{\text{LtetO}}$ module as the inducible input component and the $\text{LacI}/P_{\text{lac}}$ and CI/P_{R} modules as the first and second repressor module, respectively (Hooshangi et al., 2005). Figure 13.4B shows the experimentally measured population-averaged steady state network outputs at varying concentration of the aTc inducer when a fluorescent reporter is expressed from the P_{LtetO} (P_1), the P_{lac} (P_2), and the P_{R} (P_3) promoter, respectively. These population-averaged protein abundance curves have the correlations expected for the network. While the expression from the P_1 and P_3 promoters show a positive correlation with the input aTc concentration (that is, high-pass detection), expression from P_2 shows a negative correlation (that is, low-pass detection). When fitted to a Hill function, the Hill coefficients for the steady state response in the cascades of length one, two, and three are 2.3, 7.0, and 7.5, respectively. In other words, increased length of transcriptional regulatory cascades improves the sensitivity to the input signal by enabling more pro-

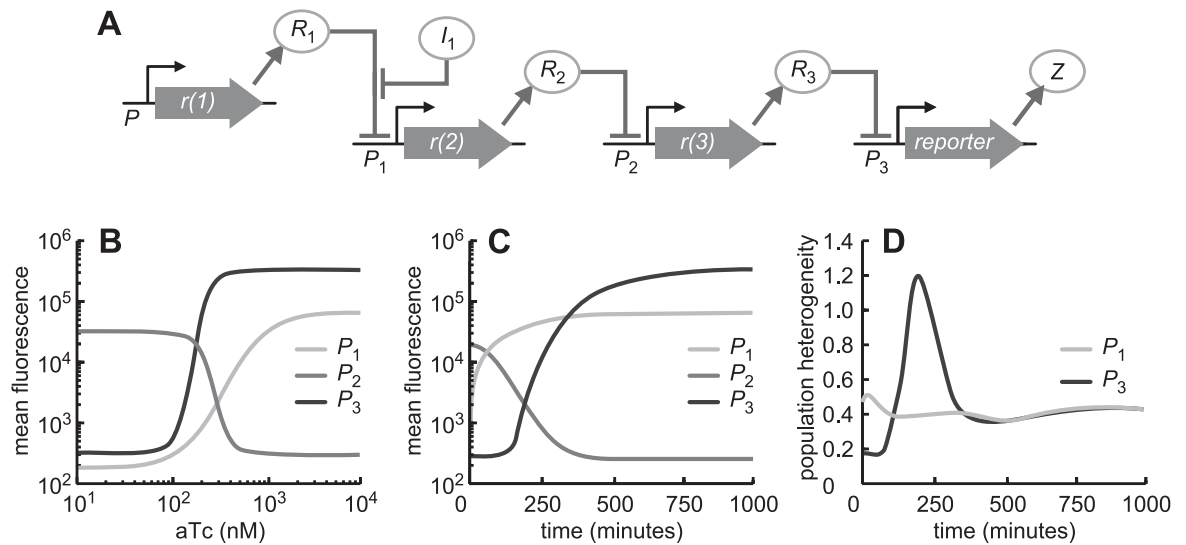


Figure 13.4 **A.** Architecture of the three-step repressor cascade. **B.** Population-averaged steady state expression levels obtained by expressing a fluorescence reporter gene at different steps in the cascade ($P_1 = P_{LtetO}$, $P_2 = P_{lac}$, $P_3 = P_R$) when the concentration of the inducer of the first transcriptional module (aTc) is varied. **C.** Time course of population-averaged expression levels at the different steps in the cascade following induction. **D.** Relative population heterogeneity (standard deviation over the mean) at steps one and three in the cascade following induction. Based on experimental data from (Hooshangi et al., 2005).

nounced all-or-nothing steady state responses. This phenomenon can also be found in naturally occurring regulatory motifs such as signal transduction phosphorylation cascades (Ferrell Jr., 1996).

It is also interesting to compare the time course of expression induction at the different steps in the cascade following aTc induction. This is done in figure 13.4C. While protein synthesis from the first promoter begins immediately after addition of aTc, there is a significant time lag in the repression and activation of the second and third promoter, respectively. The abundance of the protein expressed from P_{LtetO} (P_1) reaches the 50% of maximal abundance after ~ 15 minutes, and it takes about 200 and 300 minutes for the proteins expressed from the P_{lac} (P_2) and the P_R (P_3) promoters to pass the 50% mark. While a model based on equation 13.1 predicts such delays, the experiments give an idea of the relative time scale involved in transcriptional regulation and the response-delay introduced as the regulatory signal propagates through the network. Specifically, the cell division time for *E. coli* is typically ~ 45 –120 minutes depending on the strain and the growth conditions, meaning it may take several generations for a full transcriptional response to be realized.

Another important observation that can be deduced from the time series experiment is that the regulatory signal propagates through the cascade at very different rates in individual cells. Figure 13.4D compares the relative variability in fluorescence among cells, measured as the standard deviation over the mean, and changes following induction in the cascades of length one and three. While the cell-to-cell variability changes little as time progresses for the one-step cascade, indicative of

a fairly homogenous response, it changes significantly for the three-step cascade and reaches a peak value after about 200 minutes. At this time, which roughly corresponds to the point where expression from the P_R promoter is initiated, the cell population is highly heterogeneous. Hence, the increased steady state sensitivity in the longer cascade comes at the cost of a response that initially is highly asynchronous. An implication of this in terms of regulatory robustness is discussed further in section 13.6 in the context of feedback networks.

13.5 Feedforward Networks

Genetic feedforward networks are circuits in which transcriptional regulatory modules are configured with a common input that propagates through parallel cascades, and ultimately converge to regulate a shared downstream promoter. Several endogenous feedforward motifs have been documented (Lee et al., 2002) and three-gene networks with this architecture appear more frequently in cellular regulation than expected based on randomized networks (Shen-Orr et al., 2002). Modeling predicts that the three-gene feedforward networks support a variety of properties ranging from transcriptional response delay and filtering to the generation of transient pulses of gene expression (Mangan and Alon, 2003). Here, we limit our discussion to feedforward networks engineered in *E. coli* by interconnecting transcriptional regulatory modules in table 13.1. The first feedforward network (section 13.5.1) is composed of three genes and is designed to generate a transient pulse in response to a persistent inducing signal. The second network is composed of five genes and enables cells to respond to an inducing signal when the inducer concentration is within a specific range (section 13.5.2).

13.5.1 Pulse-Generating Network

When the downstream promoter in a feedforward network receives both an activating and a repressing signal, a transcriptional pulse can be generated if the repressing signal is delayed compared to the activating signal. Such a delay is realized if the repressing signal has to propagate through a higher number of transcriptional modules than the activating signal (see figure 13.4B). Hence, the feedforward network depicted in figure 13.5A should be able to generate a gene expression pulse. In this network, an inducing input signal (S_1) activates the transcription of a reporter gene from a multi-input promoter (P_{12}) and as well as the expression from the P_1 promoter of a repressor (R_2) of the P_{12} promoter.

Ignoring basal expression and modeling the expression from the P_{12} promoter as a product of an activating and a repressing Hill function, the feedforward network can be described by the following ODEs:

$$\begin{aligned} \frac{d[R_2]}{dt} &= \frac{k_1 \cdot s^{n_1}}{1 + s^{n_1}} - d_2 \cdot [R_2] \\ \frac{d[Z]}{dt} &= \frac{k_{12}}{1 + ([R_2]/K_2)^{n_2}} \cdot \frac{s^{n_1}}{1 + s^{n_1}} - d \cdot [Z] \end{aligned} \quad (13.6)$$

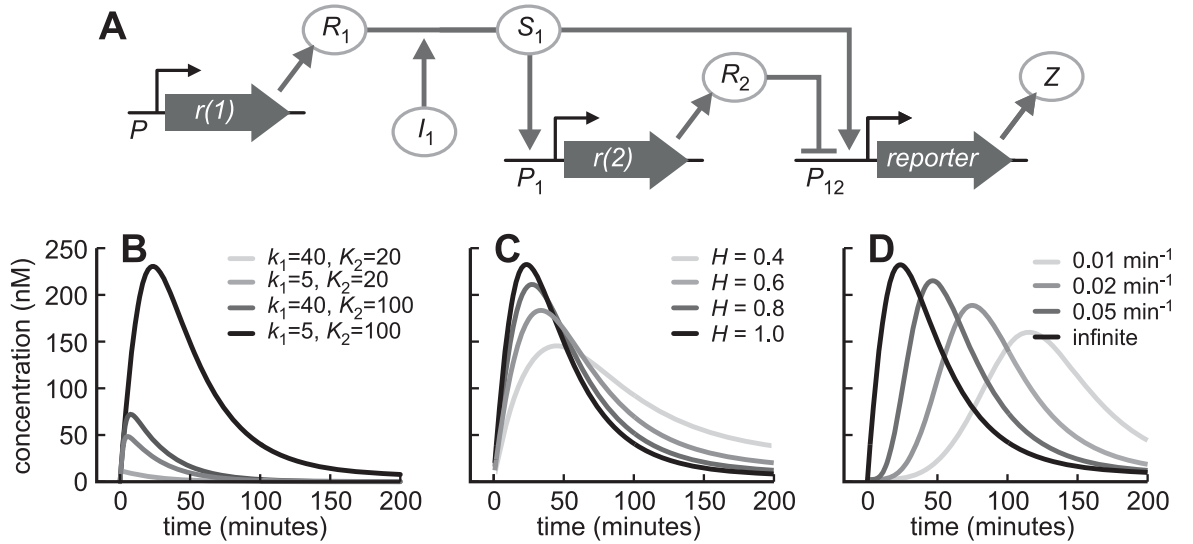


Figure 13.5 A. Architecture of the pulse-generating network. B–D. Simulations of the network model demonstrating the effect of changing parameters (panel B) at full induction ($H = 1$), the level of induction (panel C) and the rate of inducer accumulation (panel D). Unless otherwise indicated, parameter values are (in nM/min): $k_1 = 5$, $k_{12} = 20$, $d_2 = 0.01$, $d = 0.04$, (in nM): $K_1 = 1$, $K_2 = 100$, and $n_1 = n_2 = 3$. Inducer accumulation is modeled by setting $s(t) = k_s \cdot t$ with k_s being the rate of accumulation.

where $s = S_1/K_1$ is the inducing signal of P_1 (that is, dependent on the inducer concentration $[I_1]$). As before, it is not necessary to include the concentration of the R_1 protein because its concentration can be assumed constant.

Without resorting to computer simulations, let us see if we can generate intuition about the network dynamics directly from the ODEs. To do this, we define the induction level H as $H = s^{n_1}/(1 + s^{n_1})$ and find the steady states of the system. They are given by:

$$\begin{aligned} [R_2]_{ss} &= \frac{k_1}{d_2} \cdot H \\ [Z]_{ss} &= \frac{k_{12}}{d} \cdot \frac{K_2^{n_2} \cdot H}{K_2^{n_2} + [R_2]_{ss}^{n_2}} \end{aligned} \quad (13.7)$$

Let us consider the case where the induction level is constant and the lifetime of the repressor is so long that its decay can be assumed negligible. In this case, the accumulation of repressor following induction at $t = 0$ is given by $[R_2](t) = k_1 \cdot H \cdot t$. This reduces equation (13.7) to a time-dependent ODE for the output concentration that is given by:

$$\frac{d[Z]}{dt} = \frac{k_{12}}{1 + (k_1 \cdot H \cdot t / K_2)^{n_2}} - d \cdot [Z] \quad (13.8)$$

This equation captures the initial high transcription rate from P_{12} , which leads to an overshoot of the steady state in equation (13.7), and the decrease in this rate as repressor accumulates with time. It also indicates that the duration of the pulse (and hence its magnitude) is linked to the maximal rate of repressor synthesis k_1 , the Hill constant K_2 and the concentration of the induction level H . Specifically, a 50%

decrease in transcription occurs after a time period given by $t_{0.5} = K_2/(k_1 \cdot H)$. Hence, increasing K_2 or decreasing k_1 is predicted to cause a longer pulse and higher amplitude. This prediction is validated by the model simulations presented in figure 13.5B.

Suboptimal induction, that is, values of H less than one, is also predicted to increase the duration of the pulse. However, because the rate of output protein synthesis depends on the induction level, suboptimal induction is expected to decrease the amplitude. Additionally, because the length and the amplitude of the pulse depend on how fast the repressor accumulates, they should also depend on the rate at which the inducing signal accumulates. These predictions are validated by the model simulations presented in figure 13.5C and figure 13.5D, respectively.

The pulse-generating network is implemented experimentally by expressing the CI repressor from the AHL-activated P_{lux} promoter and GFP from the multi-input promoter designated P_{luxOR} . This promoter is obtained by inserting a CI binding site into the P_{lux} promoter to achieve repression of AHL-activated transcription by CI (Basu et al., 2004). The experimental observations reflect well the results of the above analysis. Figure 13.6A shows population-averaged temporal responses of four *E. coli* strains harboring networks constructed with different rates of CI synthesis (*cI*-RBS mutations) and different binding affinities of CI to the P_{luxOR} promoter (operator mutations) following induction with saturating AHL concentrations. It is seen that the effects of the mutations are in agreement with the model predictions. Due to a high repressor synthesis rate (high k_1) and strong repressor binding to the P_{luxOR} promoter (K_2), a pulse is not generated in the original network. Mutations that decrease the repressor synthesis rate or the operator binding strength yield a pulse with intermediate duration and amplitude. The best network performance is obtained when these mutations are combined.

In a second set of experiments, the temporal response was measured after induction with different AHL concentrations using the network with the best performance. As shown in figure 13.6B, at AHL concentrations below 47 nM, the pulse amplitude is decreased and its duration shortened. At an AHL concentration of 4.7 nM, the pulse can hardly be observed. In other words, the system responds differently at nonsaturating AHL concentrations as predicted by the model analysis. A third set of experiments measured the network response to different rates of AHL accumulation. The results are shown in figure 13.6C. As the rate of AHL accumulation is decreased, the onset of the pulse is delayed, and its amplitude decreased. This is also in agreement with the model prediction.

The experimental results in figure 13.6A–C are population-averaged responses obtained in a well-mixed environment. This leaves open the question of how cells harboring the feedforward network respond in an environment where signal diffusion plays an important role. Figure 13.6D illustrates the results of an experiment designed to determine the spatio-temporal response at the level of single cells. In these experiments, cells harboring the pulse-generating feedforward network are placed adjacent to *E. coli* cells that synthesize and emit AHL. The AHL-emitting “sender” cells harbor an aTc-inducible promoter controlling the expression of the

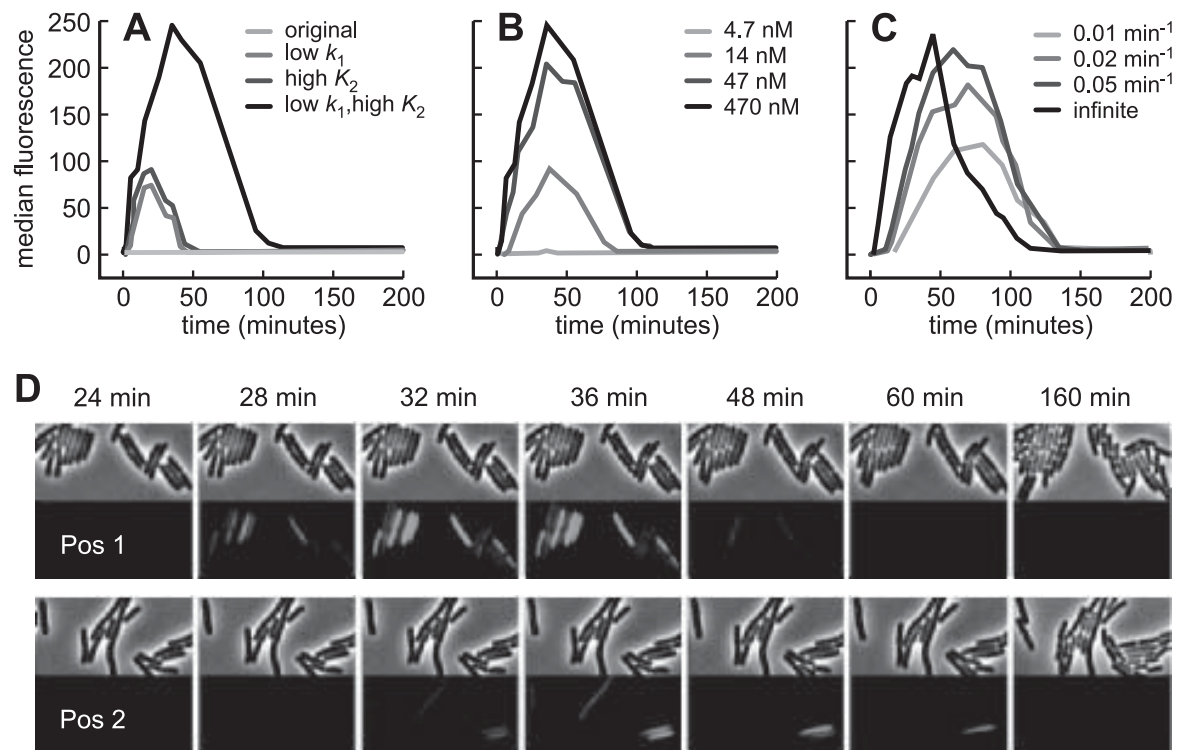


Figure 13.6 A–C. Experimental validations of the model predictions in a pulse-generating network activated by AHL. The effect of (A) introducing mutations to change the k_1 and K_2 parameters, (B) different inducer concentrations, and (C) changing the rate of inducer accumulation. Based on data from (Basu et al., 2004). D. Responses of pulse-generating cells placed at different distances from nearby sender cells to AHL synthesized by the senders. Notice that the response of cells in position 2, which is farther away from the senders, is delayed, and the maximum pulse amplitude is diminished.

enzyme (LuxI) that synthesizes AHL from common metabolites. As a result, the sender cells produce AHL when treated with aTc. The inducer subsequently diffuses into the environment and establishes an AHL concentration gradient. Figure 13.6D shows the phase-contrast and fluorescence microscopy images of “receiver” cells harboring the feedforward network taken at different time points and different distances from a colony of AHL-emitting senders. While there is distinct variability in the response from one cell to another, it is seen that single cells respond to increased AHL by generating a pulse of fluorescence. Moreover, AHL-induction elicits a response in receiver cells that depends on the distance from the senders. Because of AHL diffusion, the rate of AHL accumulation is slower farther from the AHL-emitting source. This allows receiver cells to differentiate between signals originating from nearby and distant senders.

13.5.2 Concentration Band Detection

The pulse-generating system discussed in the previous section is an example of the complex responses that can be generated from transcriptional networks combining one-step and two-step linear cascades in a feedforward architecture. In this section, we investigate a network in which a two-step and a three-step cascade are

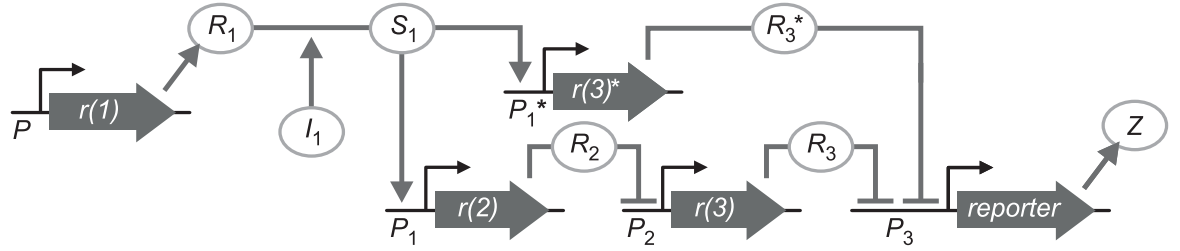


Figure 13.7 Architecture of the band detection feedforward network obtained by combining a high-pass two-step linear repressor cascade and a low-pass three-step linear repressor cascade. This combination enables the synthesis of the reporter protein only when the inducer I_1 is within a specific range.

activated by the same input and regulate the same output. The resultant five-gene feedforward system activates the expression of the output gene within a finite concentration range of the inducing signal, that is, concentration band detection, and supports the formation of spatial patterns in response to a gradient in the inducing signal. These experiments demonstrate a mechanism referred to as the “French flag model” in developmental biology (Wolpert, 2002) where cells read and respond to spatial information encoded in a “morphogen” gradient by having sharp induction thresholds.

Figure 13.7 illustrates the schematics of the five-gene feedforward network. Along the three-step branch, the inducing signal S_1 , which is generated by a combination of the regulator R_1 and its inducer I_1 , activates the expression of the R_2 , which, in turn, inhibits the expression of the R_3 repressor. In the final step, the repressor R_3 inhibits the transcription of the reporter protein. Along the two-step branch, the inducing signal activates the expression of the repressor designated R_3^* , which is functionally equivalent to the R_3 repressor (that is, it also inhibits the expression of the reporter protein).

How will the system respond to different levels of the inducing signal? To answer this question, we look at the steady state concentration of the output reporter protein. Since R_3 and R_3^* are assumed to be functionally identical, it is given by:

$$[Z]_{ss} = \frac{k_3}{d} \cdot \frac{K_3^{n_3}}{K_3^{n_3} + ([R_3]_{ss} + [R_3^*]_{ss})^{n_3}} \quad (13.9)$$

The steady state concentrations of the three repressors are given by:

$$\begin{aligned} [R_2]_{ss} &= \frac{k_1}{d_2} \frac{s^{n_1}}{1 + s^{n_1}} \\ [R_3]_{ss} &= \frac{k_2}{d_3} \frac{K_2^{n_2}}{K_2^{n_2} + [R_2]_{ss}^{n_2}} \\ [R_3^*]_{ss} &= \frac{k_1^*}{d_3} \frac{s^{n_1}}{1 + s^{n_1}} \end{aligned} \quad (13.10)$$

where k_1^* is the rate of R_3^* expression from the P_1^* promoter, and the inducing signal S_1 is expressed relative to the value that yields 50% response ($s = S_1/K_1$).

Along the two-step branch, the steady state concentration of R_3^* rises as the inducing signal increases and the expression of the output protein is inhibited when the inducing signal is high. This branch acts as a low-pass detector (see figure 13.4). The steady state concentration of R_3 in the three-step branch shows the opposite correlation and acts as a high-pass detector. For appropriately matched parameter values, this may leave a gap in the concentration of the repressors of transcription from P_3 at intermediate values of the inducing signal. It is in this gap that the output protein is synthesized.

The boundaries in the concentration of the inducing signal between which the output is expressed can be obtained from the steady states in equation 13.10 as the values S_{low} and S_{high} where R_3^* and R_3 have 50% of their maximal concentrations, respectively. They are given by:

$$S_{\text{low}} = K_1 \quad S_{\text{high}} = K_1 \sqrt[n_2]{\frac{d_2 K_2}{k_1 - d_2 K_2}} \quad (13.11)$$

Therefore, a gap in the total concentration of repressors may occur if $k_1 > 2K_2 d_2$. If a sufficient gap exists, the range of band detection can be shifted by modifying the value of K_1 .

The five-gene feedforward network is implemented experimentally (Basu et al., 2004) using the AHL-activated LuxR/ P_{lux} module to regulate the expression of the CI protein (R_2), which, in turn, regulates the expression of the LacI protein (R_3) from the P_R (P_2) promoter. The system output protein is expressed from the P_{lac} (P_3) promoter. These transcriptional regulatory modules comprise the three-step branch of the network. Along the two-step branch, AHL activates transcription of a variant of the *lacI* gene, designated *lacI*^{M1}, that differs in its DNA sequence from that of the *lacI* gene, but encodes a protein with the same amino-acid sequence. The protein product encoded by *lacI*^{M1} (R_3^*) is thus functionally identical to LacI.

The band detection network is implemented in different versions: one using the wild-type LuxR protein, designated BD2, the other, designated BD1, with a mutant variant of LuxR that is hypersensitive to AHL. In the latter, less AHL is required to achieve the same expression level from the P_{lux} promoter, corresponding to a decreased value of the Hill constant K_1 . Accordingly, the range of AHL concentrations detected by the two versions should be different, with the mutated LuxR network expressing the output protein of the system at a lower inducer concentration. Figure 13.8A and figure 13.8B show that this differential response is also observed experimentally. In figure 13.8A, the measured steady state input-response of the two-step branch is shown at varying AHL concentrations. It is seen that the AHL concentration yielding 50% response is decreased more than 10-fold when the hypersensitive LuxR variant is employed. In figure 13.8B, the observed steady state input-response of the five-gene band network is shown at varying AHL concentrations. As predicted by the model, BD1 cells activate the expression of the system output at a lower range of AHL concentration than BD2 cells.

The different ranges of AHL detected by strains harboring the different band detection networks enable multicellular pattern formation. Figure 13.8C shows the

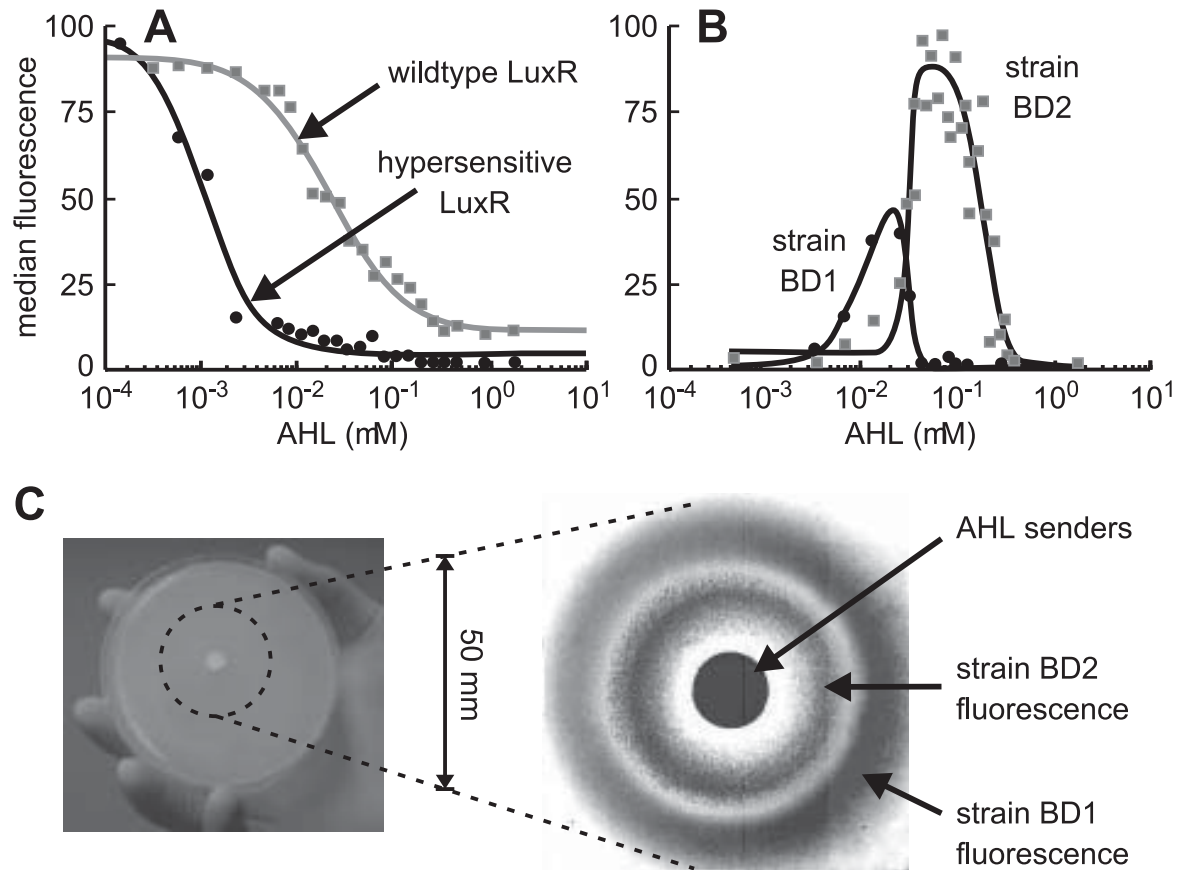


Figure 13.8 **A.** Population-averaged signal-response curves for cells harboring two versions of the two-step branches with a wildtype and a hypersensitive LuxR mutant, respectively. **B.** Signal-response curves for the five-gene feedforward network with the wildtype LuxR (strain BD2) and the mutant LuxR (strain BD1). **C.** Formation of a target pattern within a bacterial lawn containing a mixture of the BD1 and BD2 strains in the presence of AHL-emitting cells in the center of the lawn. Data from (Basu et al., 2004).

formation of a target pattern in an experiment where an AHL-emitting cell strain is grown at the center of a lawn containing a mixture of BD1 and BD2 cells. The BD2 cells turn on the expression of a fluorescent reporter gene at a short distance from the AHL-emitting cells, but remain quiescent farther away. On the other hand, the BD1 cells are quiescent near the center of the lawn and express a differently colored fluorescent reporter only at a distance from the AHL-emitting cells.

13.6 Feedback Networks

The experiments involving feedforward networks demonstrate how complex dynamics can be generated by combining linear signaling cascades. In natural regulatory systems, such behaviors are frequently generated in networks incorporating feedback loops as an additional control feature. As discussed in chapter 6, feedback control enables complex dynamics, such as bistability, hysteresis, and oscillations.

There are numerous examples of such behaviors in natural genetic circuits. For example, LacI is a key component of a natural genetic feedback network that exhibits bistability (Ozbudak et al., 2004). There are also many examples of gene regulatory feedback networks supporting dampened or sustained oscillations. They include, for example, the circadian clocks discussed in chapter 2 and chapter 12, and the Mdm2-p53 network discussed in chapter 6. A motivation for the implementation of synthetic gene regulatory feedback systems is to complement the analysis of mathematical models of natural circuits with investigations of how feedback networks behave *in vivo*. In this section, we discuss several genetic feedback networks implemented in *E. coli* to create cells capable of complex temporal dynamics and the mathematical models used to design or to understand the network properties.

13.6.1 Bistable Networks

Bistability and hysteresis are trademark features of networks that contain positive feedback or autocatalysis. Here, we investigate two single-gene positive feedback networks giving rise to hysteresis (Atkinson et al., 2003) and bimodal population distributions (Isaacs et al., 2003), respectively, and a two-gene system designed to operate as a bistable genetic toggle switch (Gardner et al., 2000; Kobayashi et al., 2004). In the single-gene positive feedback system depicted in figure 13.9A, a transcription activator R_1 binds to its own promoter and increases the rate of its own synthesis. This network can be described by the ODE:

$$\frac{d[R_1]}{dt} = a_1 \cdot k_1 + \gamma \frac{k_1 [R_1]^{n_1}}{K^{n_1} + [R_1]^{n_1}} - d_1 \cdot [R_1] \quad (13.12)$$

where the parameter γ is a measure of the feedback control strength. Because at this point we are interested in using the model to reveal general trends, it is useful for the analysis to introduce new dependent variables to reduce the number of unknown parameters. For equation 13.12, a useful normalization is to use the dimensionless concentration r_1 defined by $r_1 = [R_1]/K_1$ and dimensionless time τ defined by $\tau = d_1 \cdot t$. This corresponds to expressing the protein concentration relative to that yielding 50% response and time relative to the protein lifetime, regardless of the actual value of these parameters. Using the chain rule, the normalized form of equation 13.12 is obtained as

$$\frac{dr_1}{d\tau} = a \cdot \kappa_1 + \gamma \frac{\kappa_1 \cdot r_1^{n_1}}{1 + r_1^{n_1}} - r_1 \quad (13.13)$$

where κ_1 is defined by $\kappa_1 = k_1/K_1/d_1$. Similarly normalized equations will be used in the remaining sections of this chapter.

Figure 13.9B shows a bifurcation diagram obtained by plotting the steady state solutions of equation (13.13) as a function of the feedback control strength. The steady state curve has the “S”-shape characteristic of bistable systems. At low feedback strength, there is little or no activation, and expression occurs essentially at basal levels. At high feedback strength, the promoter is more or less fully activated

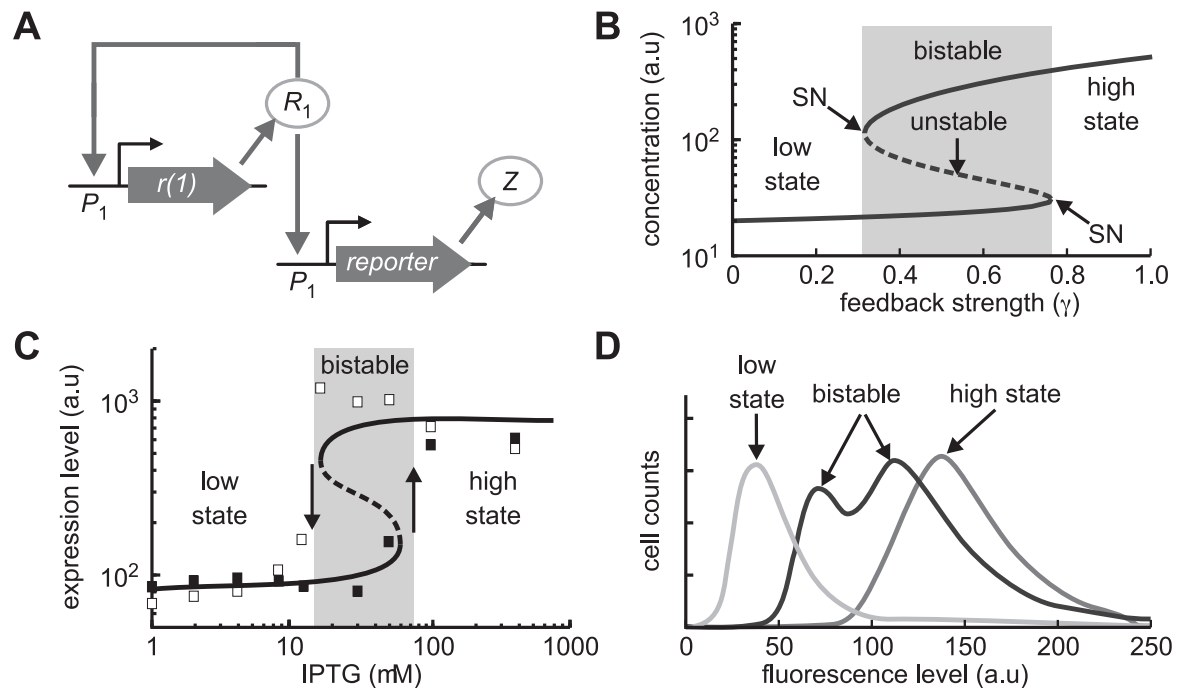


Figure 13.9 Bistability in single-gene autocatalytic networks. **A.** Network architecture. **B.** Bifurcation diagrams for the network model. Full and broken curves indicate stable and unstable steady states, respectively. The saddle-node (SN) bifurcations are located where the stable and unstable steady states collide. Parameter values are: $a = 0.1$, $\kappa = 5$, and $n = 3$. **C.** Bistability and hysteresis observed in a NtrC positive feedback network when the strength of NtrC-activated transcription is varied with IPTG. Closed and open squares correspond to cells initially in the low and high states, respectively. Based on data from (Atkinson et al., 2003). **D.** Transitions between uni- and bimodal population distributions observed in a CI/ P_{RM} feedback network corresponding to different strengths of the feedback loop. Based on data from (Isaacs et al., 2003).

and expression takes place at a rate close to maximal. These two states co-exist at intermediate values of the feedback control strength parameter with the region of bistability demarcated by two saddle-node bifurcations located at a value of γ slightly above 0.3 and just shy of 0.8.

Several synthetic single-gene autocatalytic gene networks have been constructed with the purpose of generating bistability (Becskei et al., 2001; Isaacs et al., 2003; Atkinson et al., 2003). One system (Atkinson et al., 2003) is constructed such that the transcription factor NtrC activates its own expression from a modified NtrC-responsive P_{glnA} promoter and that of a reporter gene from the promoter of the *glnK* gene (P_{glnK}). The modified P_{glnA} promoter is engineered such that the ability of NtrC to activate transcription is attenuated by LacI. This is achieved by inserting LacI binding sites such that the repressor competes with the activator for promoter access. This allows for an indirect means of modulating the feedback control strength using IPTG. In cells that express LacI at high levels, increasing the IPTG concentration enables more efficient activation by NtrC of transcription from the modified P_{glnA} promoter.

Figure 13.9C shows the experimentally observed effect of varying the feedback control strength on the population-averaged expression of cells harboring the IPTG-

sensitive, NtrC positive feedback system. They are in excellent agreement with the model predictions. When grown in the absence of IPTG (that is, low feedback strength), cells express the reporter protein at low levels because of efficient repression by LacI. When these cells are exposed to increased inducer concentrations (closed squares in figure 13.9C), the measured signal-response curve is fairly flat below a critical IPTG concentration where the reporter expression changes sharply from low to high levels. On the other hand, when cells initially grown with high IPTG concentrations and fully activated (open squares) are exposed to decreased concentrations of IPTG, expression levels remain high until a critical concentration where a sharp transition to low expression is observed. At identical intermediate IPTG concentrations, corresponding to intermediate strength of the feedback control, cell populations adopt a high or a low expression state depending on the initial conditions. Hence, the network endows cells with the ability to support bistability and hysteresis.

Another single-gene positive feedback system engineered to display bistability (Isaacs et al., 2003) employs the CI-activated P_{RM} promoter to control the expression of a mutated cI gene (designated $cI857$) encoding a temperature-sensitive variant of the CI protein. The CI protein also activates the transcription of a GFP-encoding gene allowing the measurement of gene expression at the level of single cells. The temperature-dependent activity of the $cI857$ -encoded CI protein enables modulation of the feedback strength through temperature variation. The activity of the CI variant decreases with increased temperature. Hence, a low temperature corresponds to a high feedback strength and high temperature corresponds to a low value of this parameter. The model thus predicts a low expression state at high temperature, a high expression state at low temperature, and bistability at intermediate temperatures. Figure 13.9D illustrates the population-distribution of fluorescence from cells harboring the CI/ P_{RM} feedback network at three different temperatures. For low and high temperatures, the population distributions contain a single peak and cells are in a high state when they are grown at low temperature and in a low expression state when they are grown at high temperature, respectively. At the intermediate temperature, the population-distribution is bimodal, indicating that cells transition frequently between the low and the high expression states due to noise-induced transitions. A detailed model of the circuit where the deterministic equations are augmented with stochastic terms accounts well for the observed distributions (Isaacs et al., 2003).

The toggle switch network, which is illustrated in figure 13.10A, is an example of a two-gene system designed and implemented to allow *E. coli* cells to be switched between two distinct expression states in response to external stimuli. This system, which represents a multi-component motif (Lee et al., 2002) with indirect positive feedback, is composed of two genes encoding transcription factor proteins, R_1 and R_2 , that inhibit each other's expression. Because of this mutual repression, the network can be either in a state with high R_1 expression and repressed R_2 transcription, or in a state with high R_2 expression and repressed R_1 transcription.

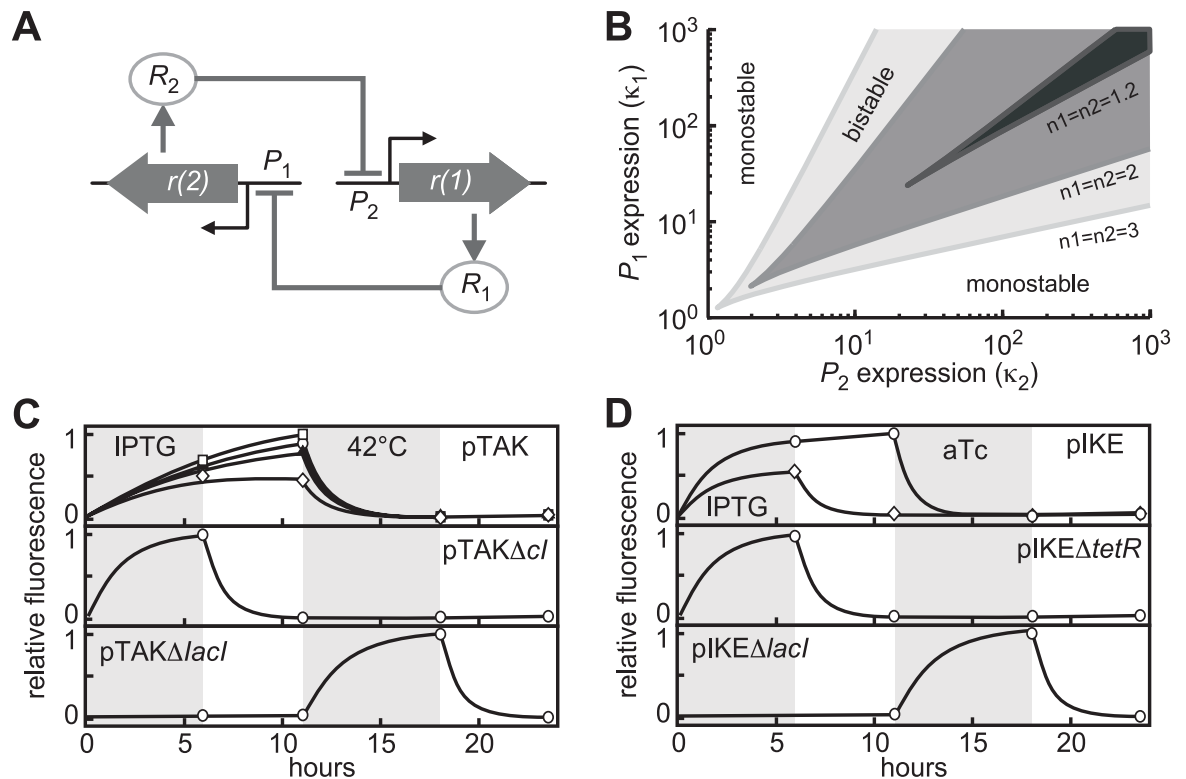


Figure 13.10 The bistable toggle switch. **A**. Architecture of the network. **B**. Two-dimensional bifurcation diagrams indicating the boundaries between bistable and monostable regions in the κ_1 , κ_2 parameter space at different values of the Hill coefficients. **C–D**. Transitions between high and low expression states in the pTAK and the pIKE toggle switch networks, respectively. The networks, pTAK Δcl , pTAK $\Delta lacI$, pIKE $\Delta tetR$, and pTAK $\Delta lacI$, are controls in which one of the repressor genes is eliminated. Based on data from (Gardner et al., 2000).

The experimental implementation of the toggle switch network is guided by the analysis of the dimensionless ODEs (Gardner et al., 2000):

$$\frac{dr_1}{d\tau} = \frac{\kappa_2}{1 + r_2^{n_2}} - r_1, \quad \frac{dr_2}{d\tau} = \frac{\kappa_1}{1 + r_1^{n_1}} - r_2 \quad (13.14)$$

where the repressor concentrations are expressed relative to the appropriate Hill constant and time relative to the protein lifetime (which is assumed to be the same for the two repressors). Conditions that make bistability more likely are high maximal expression levels (that is, high values of κ_1 and κ_2) and high Hill coefficients. This can be seen from the bifurcation diagrams in figure 13.10B, which show the location of saddle-node bifurcations, that is, the boundaries between mono- and bistability, in the κ_1 , κ_2 parameter plane for different values of the Hill coefficients. Increased values of the Hill coefficients enlarge the bistable region in the κ_1 , κ_2 parameter space, and an increased value of one of the maximal expression rates allows for bistability in a wider range of values of the other.

The toggle switch network is implemented experimentally using different transcriptional regulatory modules. Gardner et al. (2000) constructed two versions; one employs the LacI/ P_{trc} and the TetR/ P_{LtetO} modules and is designated as pIKE.

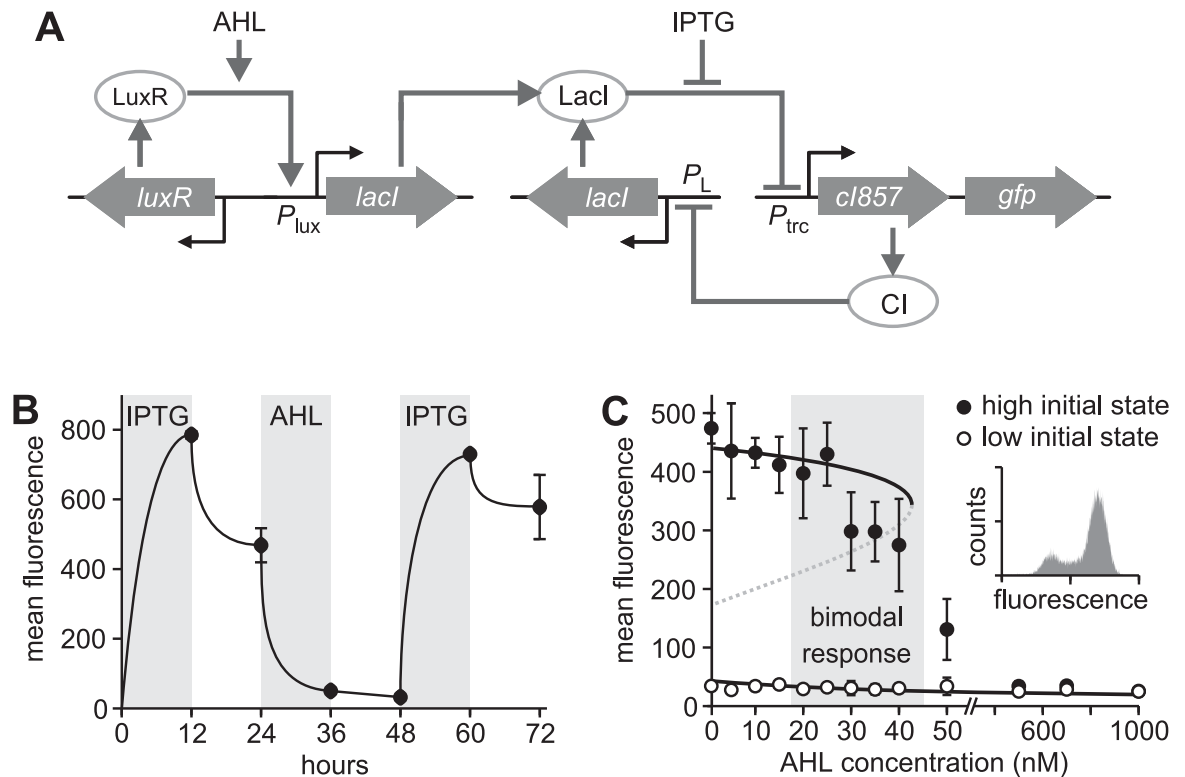


Figure 13.11 Extending the toggle switch network. **A**. The pTAK toggle switch is augmented with the AHL-activated LuxR/ P_{lux} to convert the network into a AHL sensor. **B**. Flipping between stable expression state by application of AHL to activate LacI expression or IPTG to activate CI expression. **C**. Hysteresis is observed when cells initially in the low (open circles) or high (closed circles) expression states are exposed to AHL at varying concentrations. Based on data from (Kobayashi et al., 2004).

The other employs the LacI/ P_{trc} and the temperature-sensitive CI/ P_L modules and is designated as pTAK. The expression state is monitored by co-expressing GFP with the *cl857* gene (pTAK) or the *tetR* gene (pIKE). Two variants of the pIKE network with strong and weak *lacI*-RBS sequences demonstrate that the maximal rate of expression, in agreement with model predictions, is an important parameter for the emergence of bistability. Figure 13.10B shows the effect of treating cells in the high LacI state with IPTG to activate TetR expression. Both variants respond by expressing the reporter protein. When IPTG is removed, cells harboring the network with the weaker *lacI*-RBS maintain the high expression state while those harboring the variant with the stronger *lacI*-RBS revert to the low expression state. The cells that remain in the high TetR state require addition of aTc to reactivate LacI expression. Hence, only the network with the weak *lacI*-RBS supports bistability. Four variants of the pTAK network, also with different RBS sequences, all exhibit bistability. This is shown in figure 13.10C. Addition of IPTG to inhibit LacI induces a transition to a high expression state (low LacI/high CI), and the latter is maintained when IPTG inducer is removed. A subsequently applied transient temperature increase (to deactivate CI) induces a transition back to the low expression state (high LacI/low CI).

In an extension to the toggle switch (Kobayashi et al., 2004), mathematical modeling is used to guide the experimental implementation of “programmable” cells in which the pTAK toggle responds to signals from other gene regulatory networks. In one of the implementations, the LuxR/ P_{lux} module is used to drive additional synthesis of LacI. The resultant five-gene network is depicted in figure 13.11A. In this system, the toggle switch can be flipped back and forth by adding AHL to increase the LacI synthesis rate and IPTG to increase the CI synthesis rate. This is illustrated in figure 13.11B, which shows population-averaged expression levels following induction with AHL and IPTG. Cells initially switch to the high expression state (high CI) following IPTG treatment and remain in this state when IPTG is removed. A transition to the low expression state (high LacI) occurs when these cells subsequently are treated with AHL, and the low state is maintained when AHL is removed. These cells are still responsive, and a second treatment with IPTG induces a transition to the high expression state.

The system in figure 13.11A also supports hysteresis. Figure 13.11D shows the result of an experiment where cells that were initially prepared in the high CI state (high fluorescence) or the high LacI state (low fluorescence), respectively, are exposed to AHL at varying concentrations. The high LacI state is, as expected, unaffected by AHL, and cells remain in the low fluorescence state regardless of the AHL concentration. However, the high CI state is sensitive to AHL. At inducer concentrations less than 20 nM, cells remain in the high expression state. On the other hand, at inducer concentrations higher than 40 nM, all the cells have switched to the low expression state. At intermediate inducer concentrations, the cell population contains a mixture of cells in high and low expression states. This bimodal response presumably arises from a combination of differences in induction threshold and noise-induced transitions, which are more likely to occur when the system is closer to the saddle-node bifurcation.

13.6.2 Oscillatory Networks

The discussion in the previous section provides examples of two complex properties, bistability and hysteresis, supported by genetic networks incorporating feedback regulation. Other complex behaviors that arise in feedback control systems are dampened and sustained oscillations. An example of a synthetic gene regulatory system capable of generating oscillations is obtained by adding a negative feedback to a three-step linear repressor cascade. The resultant system, which is referred to as the Repressilator (Elowitz and Leibler, 2000), is illustrated schematically in figure 13.12A. In the network, repressor R_1 inhibits the expression of repressor R_2 , repressor R_2 inhibits the expression of repressor R_3 , and repressor R_3 inhibits the expression of repressor R_1 .

The implementation of the Repressilator network is based on the analysis of a model describing the dynamics of repressor mRNA and protein concentrations. The concentration of mRNA is included explicitly because the separation of transcription and translation contributes to a response delay that is important for the

emergence of oscillations. Hence, equation 13.3 is used as the basis of the network model rather than equation 13.1, which, as described in section 13.3, assumes that the mRNA concentration is always in a steady state. To ease the analysis, it is assumed that each transcriptional module is characterized by the same set of parameters and that the rate constant associated with translation is equal to that associated with repressor decay. With these assumptions, the network is described by the following dimensionless equations (Elowitz and Leibler, 2000):

$$\frac{dm_i}{d\tau} = a\kappa + \frac{\kappa}{1 + r_j^n} - m_i, \quad \frac{dr_i}{d\tau} = \varepsilon(m_i - r_i) \quad (13.15)$$

where m_i and r_i represents the concentration of repressor R_i mRNA and protein, respectively, and r_j the concentration of repressor R_j regulating the expression of repressor R_i . The parameter ε is proportional to the ratio of the mRNA and the protein lifetimes.

What makes sustained oscillations possible in the Repressilator network? The answer to this question can be obtained by considering the dynamics of the linear three-step repressor network in section 13.4.4. Recall that in the linear network, the transcription of the R_1 repressor leads to increased expression from the P_3 promoter after a time delay. In the Repressilator, the R_1 repressor is expressed from the P_3 promoter. Hence, the network can be viewed as a negative feedback system with time delay (transcription of R_1 from the P_3 promoter eventually causes down-regulation of its own expression). The time delay in repression allows for the accumulation of protein product beyond the steady state level and, when the repression kicks in, for the subsequent decay in protein concentration. The mechanism causing oscillations in the Repressilator network is thus somewhat analogous to that leading to circadian clock oscillations as discussed in chapter 2 and chapter 12.

The conditions that make oscillations more likely to occur are high Hill coefficients, low basal expression levels, and short protein lifetimes. This is illustrated by the bifurcation diagrams in figure 13.12B, which show the location of the Hopf-bifurcations in the κ, ε parameter space that separate regions of oscillatory and steady state dynamics for different sets of parameter values. When the Hill coefficients are low and basal expression high, oscillations occur at intermediate values of κ when ε is greater than a critical value. In this case, the protein lifetime needs to be comparable to that of mRNA in order for oscillations to occur. Decreasing the basal expressions rate and increasing the Hill coefficients relax this requirement and allow oscillations to occur for a broader range of κ and ε values.

The Repressilator network is implemented experimentally by interconnecting the LacI/ P_{LacO} , the CI/ P_R , and the TetR/ P_{LtetO} modules such that LacI (R_1) is expressed from P_R (P_3), TetR (R_2) is expressed from P_{LlacO} (P_1), and CI (R_3) is expressed from P_{LtetO} (P_2). To decrease the lifetime of the repressor proteins, the repressor genes are “tagged” with a DNA sequence that targets the expressed protein for degradation. Figure 13.12C illustrates the temporal oscillations in fluorescence of a single cell measured when a fluorescent protein is expressed from a second

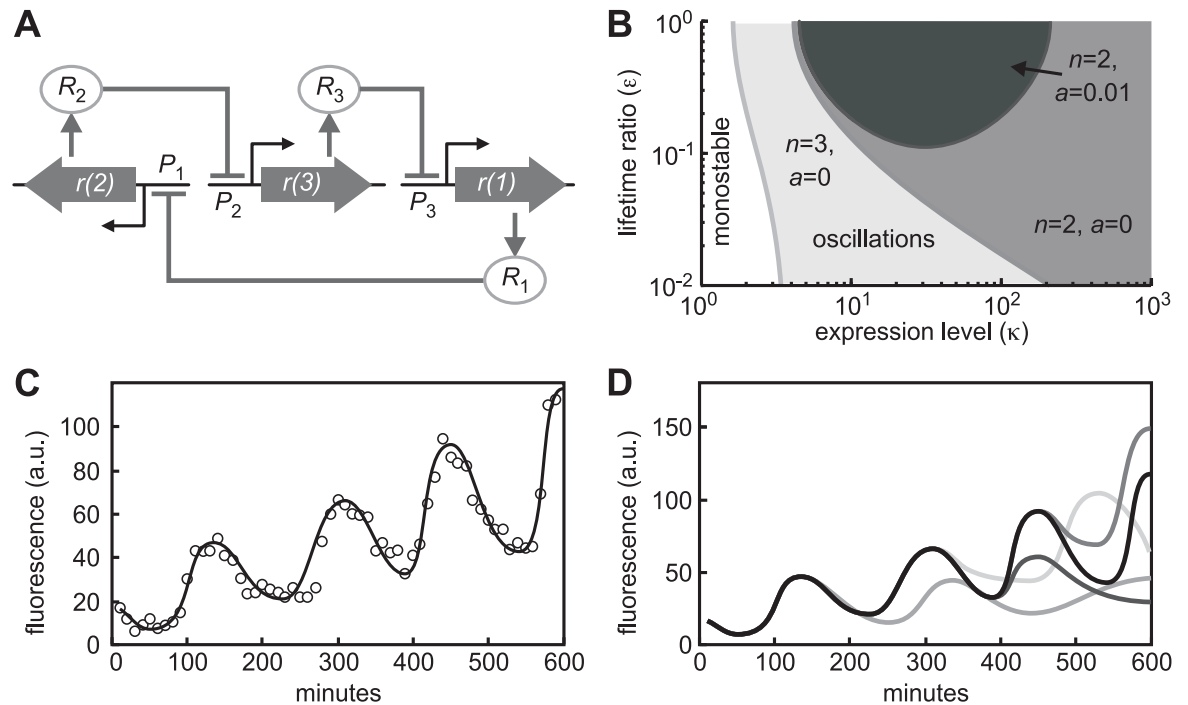


Figure 13.12 The Repressilator. **A**. Architecture of the three-gene network. **B**. Bifurcation diagram of the Repressilator model showing the regions of monostability and oscillations in the κ , ε parameter plane for different values of a and n . **C**. Time series of fluorescence emitted by a single cell harboring the Repressilator network. **D**. Comparison of the single-cell time series in **C** with those generated by its siblings. Based on experimental data from (Elowitz and Leibler, 2000).

P_{LtetO} promoter. At least 40% of cells oscillate with a period of 160 ± 40 minutes. This period is significantly longer than the average cell division time. Hence, an oscillation initiated in a mother cell is completed in a daughter cell, and the oscillation phase is passed down from one generation to the next. However, as shown in figure 13.12D, siblings display marked differences in their progression through the oscillation cycle, and the network fails to support coherent oscillations at the population level.

The Repressilator network fails to support coherent oscillations partly because of the significant differences in the rate at which the regulatory signal propagates through the three regulatory steps. Recall from section 13.4.4 that cells harboring the three-step linear cascade show marked variability in the onset of expression from the P_3 promoter. This is expected to translate directly into significant differences in the oscillation period as the differences in response times in the linear cascade are equivalent to differences in delay times in the feedback network. Hence, the number of steps in the network makes it especially susceptible to stochastic effects.

There are network architectures that generate oscillations with increased robustness against stochastic effects (Vilar et al., 2002). One example is a design, illustrated in figure 13.13A, where a transcriptional activator R_1 enhances its own expression from the multi-input promoter P_{12} and that of a repressor R_2 from the P_1 promoter. The R_2 repressor in turn attenuates the transcription of the activator

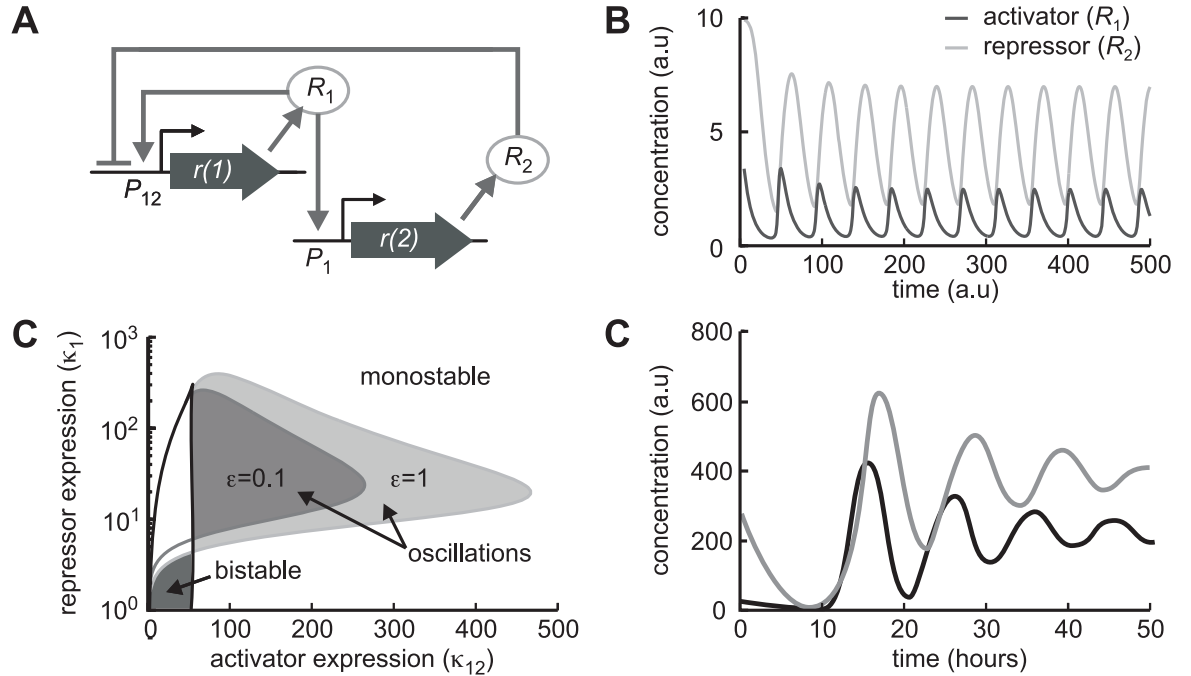


Figure 13.13 Mixed positive and negative feedback oscillator. **A.** Architecture of the network. **B.** Example of oscillations generated in simulations of the network model with parameter values $\kappa_1 = 10$, $\kappa_{12} = 100$, $n = 3$, $a = 10^{-3}$, and $\varepsilon = 0.1$. **C.** Bifurcation diagram showing the regions of monostability, bistability and oscillations in the κ_{12} , κ_1 plane for different values of ε for $a = 10^{-3}$ and $n = 3$. **D.** Coherent dampened oscillations observed in cell populations carrying two variants of the network differing approximately four-fold in the level of expression from the P_{12} promoter. Based on experimental data from (Atkinson et al., 2003).

by binding to the P_{12} promoter. Using the same assumptions as in the model of the Repressilator, the network can be modeled by the following dimensionless ODEs:

$$\begin{aligned} \frac{dm_1}{dt} &= a \cdot \kappa_{12} + \frac{\kappa_{12}}{1 + r_2^n} \cdot \frac{r_1^n}{1 + r_1^n} - m_1 \\ \frac{dm_2}{dt} &= a \cdot \kappa_1 + \frac{\kappa_1 \cdot r_1^n}{1 + r_1^n} - m_2 \\ \frac{dr_i}{d\tau} &= \varepsilon(m_i - r_i) \quad i = 1, 2 \end{aligned} \quad (13.16)$$

A simulation of the system for parameters yielding sustained oscillations is given in figure 13.13B. They arise because of a time delay between the activation and the repression of transcription from the P_{12} promoter. When the repressor concentration is initially low, the positive feedback causes an increase in both activator and repressor expression. However, because the negative feedback involves an additional step, the expression of the repressor is delayed. As a result, and akin to the pulse-generating network discussed in section 13.5.1, the activator accumulates to high levels before the repressor can reach a concentration that is sufficiently high to shut down the expression from the P_{12} promoter. Once this occurs, activator expression ceases and the activator concentration declines. This in turn causes the

rate of repressor expression to decrease. The system subsequently returns to the low repressor state and a new cycle is initiated. Figure 13.13C shows the bifurcation diagram indicating the regions of monostability, bistability, and oscillations in the κ_1, κ_{12} parameter space. The bistability is associated with the positive feedback and is more likely to occur in the absence of the negative feedback (that is, for $\kappa_1 = 0$).

The network illustrated in figure 13.13A is implemented experimentally by augmenting the bistable NtrC single-gene positive feedback network discussed in section 13.6.1 with a negative feedback (Atkinson et al., 2003). Recall that in this network the strength of the positive NtrC feedback is dependent on LacI as the modified NtrC-activated P_{glnA} promoter (P_{12} in figure 13.13A) contains lacO operators. Hence, a negative feedback is readily added to the network by expressing LacI from a second NtrC-activated promoter (P_1 in figure 13.13A). Figure 13.13D shows two time series of the population-averaged expression level in cells harboring two different variants of the oscillator network that have an approximately four-fold difference in the expression rate from the modified P_{glnA} promoter. In both cases, the population-averaged expression level exhibit dampened oscillations with a period of 10–12 hours. The dampening of the oscillations is not due to cells becoming desynchronized. Measurements of the expression of a fluorescence reporter protein in individual cells (data not shown, see (Atkinson et al., 2003)) indicate that the dampening also occurs at the level of single cells. Given that the cells divide about once per hour, the coherence of the oscillations, which appears to be maintained for the duration of the experiments, that is, 40–50 generations, is quite remarkable. It confirms the prediction that a network architecture combining positive and negative feedback should be more robust against noise.

13.7 Conclusions

In this chapter, we have discussed selected synthetic gene regulatory systems implemented experimentally in *E. coli* to investigate the dynamics of transcriptional regulatory networks *in vivo* and to create strains with novel characteristics. These systems support a range of non-trivial behaviors such as cellular memory, pulse generation, spatial pattern formation, and oscillatory gene expression. In all the examples, the networks are designed with the aid of mathematical models based on fairly simple, phenomenological descriptions of relationships between input and output signals. These models are used to predict systems properties and how changes in DNA sequence affect performance and dynamics. In all cases, an excellent agreement between model predictions and experimental results is obtained. This demonstrates the close link between the current modeling methodologies and biological reality. Consequently, there is good reason to believe these methodologies are also useful for the analysis of the more complex regulatory systems found in nature. Indeed, the systems presented and analyzed elsewhere in this book strongly indicate that this is the case.

