Design and Construction of Genetic Applets

Timothy S. Gardner & James J. Collins
Center for BioDynamics &
Department of Biomedical Engineering
Boston University

August 31, 1999

Abstract: Advances in molecular biology and medicine are making gene therapy an increasingly practical method for the treatment of chronic diseases. Regulated expression of the introduced genes is necessary for safe and effective gene therapy. However, the tools available for controlling the expression of artificially introduced genes are limited. In this work, we are developing a theory and an experimental protocol for constructing artificial gene networks that can regulate temporal expression of multiple genes. Such a network, which we have termed a "genetic applet", provides the target cell with the ability to monitor and respond appropriately to changes in its environment. In effect, the cell is "reprogrammed" to perform a desired function. In addition to gene therapy, genetic applets may have applications in chemical and biological warfare defense and biological research, and may provide a new framework understanding the regulation of gene expression.

A Contents

B List of Figures and Tables

Figures

Tables

C Statement of Objectives

Our objective is to develop a theoretical and experimental framework for the design and construction of artificial gene networks. These networks, termed "genetic applets," will be designed using a combination of deterministic (*macroscopic*) and stochastic (*microscopic*) modeling techniques. A genetic applet can be programmed into DNA and delivered into cells to execute coordinated sequences of cellular functions while responding to intracellular and extracellular signals. Specifically, we plan to: (1) develop a theory of artificial gene expression and regulation using deterministic and stochastic mathematical modeling, (2) validate and refine the modeling approach by testing our predictions against experimental data for single-cell gene expression, and (3) design and construct the basic components of a genetic applet: a genetic toggle switch, an adjustable-threshold genetic switch and a two-state genetic oscillator.

D Statement of the Approach

In this project, we plan first to construct simple gene expression systems for evaluating and refining our theoretical approach. Using the Green Fluorescent Protein (GFP) as a reporter gene in *Escherichia coli*, we will collect single-cell expression data for various promoters under the regulation of their associated repressors. The single-cell data will be collected using automated image processing of digital images obtained from an epi-fluorescence microscope. This data will allow us to evaluate both the macroscopic and microscopic predictions of the model. These simple expression systems will be constructed from the individual genetic components to be used in the toggle switch. Thus, they will simultaneously serve as the stepping-stone for the construction of the toggle switch. After construction of the toggle switch, and the analysis and refinement of its function, we will construct the adjustable-threshold switch and, finally, the two-state genetic oscillator. This project will involve the collaboration of two groups at Boston University: the Center for BioDynamics and the Center for Advanced Biotechnology.

E Statement of Significance

This project represents a departure from traditional genetic engineering in that we are constructing artificial *networks* of genes. These networks may have immediate value in clinical therapies, biomedical research, and biotechnology applications. Moreover, these networks may provide a vehicle for evaluating the predictive potential of theoretical descriptions of gene networks. Because all parts of our networks

---

Abbreviations: A_{660}, Optical Density at 660 nm; CAT, chloramphenicol acetyl transferase; CCD, Charge Coupled Device; CDC, Cell Division Cycle; FRET, Free Resonance Energy Transfer; GFP, Green Fluorescent Protein; IPTG, isopropyl β-D-thiogalactopyranoside; LacZ, β-galactosidase; LB, Luria-Bertani; mRNA, messenger RNA; RBS, Ribosome Binding Site; RNAP, RNA Polymerase; SD, Shine-Dalgarno sequence; tRNA, transfer RNA.

can be designed and manipulated, they permit each aspect of the theory to be tested and refined. The improved theory can then be used to rapidly design more complex gene networks with greater functionality. Furthermore, the theory can be extended to the analysis and prediction of the function of natural gene networks.

F Key Words/Phrases

Gene regulatory networks, genetic engineering, nonlinear dynamics, statistical physics, genetic toggle switch, genetic oscillator

G Body of the Paper

G.1 Introduction and Background

Advances in molecular biology and medicine are making the manipulation of cellular function increasingly reliable and effective. These manipulations have proven to be extremely powerful in basic biological research and have even shown promise in clinical gene therapies [1]. However, the tools available for controlling the expression of artificially introduced genes are limited to simple induction schemes [2–9]. Expression systems providing temporal, sequential and, most importantly, autoregulated control of gene expression do not exist. Our objective is to dramatically extend the reach of gene-expression technology by designing and building "genetic applets". A genetic applet is essentially an artificial virus that, once delivered into a cell, would coordinate the sequential expression of multiple genes. These genes could be used to execute a series of specific tasks that would repair, enhance or modify cell function. In effect, a genetic applet, encoded into DNA, would "reprogram" the function of the target cell.

A functional genetic applet would be composed of two to tens of interacting genes. Like a computer, an automobile or any other complex machine, a genetic applet would be composed from many independent, simpler parts. A computer, for example, is constructed from resistors, capacitors, inductors, and transistors. These components are then combined into memory units, logic gates, amplifiers and other devices that are integrated to form a functional computer. Analogously, a genetic applet is constructed from genes, promoters, and proteins that are combined to form switches, memory units, sensors and other simple elements. These elements are then integrated to form a functional applet.

In order to realize our ultimate goal of constructing a genetic applet, we must first design, build and characterize the simpler components. Moreover, we must evaluate and calibrate the modeling techniques that form the basis for the genetic applet designs. Thus, our initial research will focus on building three successively more complex elements—a genetic toggle switch, an adjustable-threshold switch, and a two-state genetic oscillator—and refining the modeling techniques. The toggle switch forms the building block for the adjustable-threshold switch; the adjustable-threshold switch is the central component of the two-state oscillator; and, though not described in this document, three or more copies of the two-state oscillator can be used to construct a multi-state oscillator. A multi-state oscillator, which sequentially expresses multiple genes, forms the core of a genetic applet.

Although we have described the toggle switch, the adjustable-threshold switch and the two-state oscillator as building blocks for a genetic applet, they are, by themselves, small genetic applets. They can respond to internal and external cues, can be auto-regulated and have tunable temporal characteristics. Moreover, these devices could have immediate value in both clinical and biotechnology applications. They might, for example, be used as sensitive *in vivo* detectors of chemical or biological warfare agents, allow precisely controlled gene expression in gene therapies, or provide a mechanism for initiating apoptosis in engineered micro-organisms. These and other potential applications are discussed in greater detail in Section G.6.

Outside of their potential value in clinical treatments and biotechnology, genetic applets and their constituent parts have great value in the theoretical study of the regulation of gene expression. Gene expression in cells is an enormously complex process composed of thousands of interacting genes that coordinate everything from embryonic development to cell differentiation to immune responses. Advances in molecular biology and bioinformatics have revealed many of these regulatory genes, but they have not clarified how these genes interact to produce the observed cellular behaviors.

One theoretical approach toward understanding gene expression is to construct highly detailed models of well-characterized genetic systems such as λ *phage* [10, 11]. Another approach is to build more abstract but simpler models which characterize many of the basic features of gene regulatory networks [12]. In such efforts, researchers use modeling and simulation techniques to *reverse* engineer existing biological systems. It is hoped that the lessons learned from these models can be used to design and/or manipulate other genetic systems. Although these approaches have been moderately successful in modeling gene expression, they have yielded few practical methods for manipulating cellular function.

Genetic applets provide the basis for a *forward* engineering approach toward the study of gene expression. A forward engineering approach may be more effective than the reverse engineering approach because it permits the complete manipulation of all elements in the system. The enormous complexities of natural gene networks can be engineered out of the applets. Thus, they serve as highly simplified, highly controlled models of natural gene networks. The molecular mechanisms and interactions governing the behavior of the applets can be used to test and refine a more general theory of gene regulation. This theory can then be extended to natural gene networks.

In the past, such complex engineering of cellular function was not possible because experimental techniques were inadequate, biological data were limited, and the principles of cellular control were not well-understood. Although construction of genetic applets may still seem far-fetched, we are confident in our approach. In recent years, advances in experimental techniques have dramatically extended our ability to manipulate DNA, genes, and proteins. At the same time, large-scale genome-sequencing projects and advances in bioinformatics have created a wealth of rapidly accessible biological data. Moreover, the rapidly growing field of functional genomics has begun to produce abundant and detailed information about protein and DNA interactions. Finally, our approach represents a significant departure from traditional genetic engineering in that we do not begin with experiments. We begin by developing a theory of genetic applets using techniques from nonlinear dynamics, neural network theory, and statistical physics. We then use this theory to guide the

experimental development of these devices; and conversely, the experimental data will be used to improve and generalize the modeling techniques. To demonstrate the feasibility of our combined theoretical/experimental approach, we plan to build the following gene network systems: (1) a genetic toggle switch (i.e., a bistable switch), (2) an adjustable-threshold genetic switch, and (3) a two-state genetic oscillator.

In the following discussion, we describe the application of a combined deterministic and stochastic modeling approach to the design and construction of genetic applets. We begin, in Section G.2, with an introduction of the theoretical concepts used to model and design gene networks. In Section G.3, we describe the genetic toggle switch using a deterministic, and more intuitive, modeling approach. We then describe how analytical stochastic modeling, in combination with experimental data on the statistics of single-cell expression, will be used to construct a toggle switch with reliable and predictable macroscopic behavior. Next, we briefly describe extensions of the toggle switch to an adjustable-threshold switch and a two-state oscillator. In Section G.4, we discuss the practical issues of implementing the gene networks and collecting macroscopic, statistical and dynamic gene-expression data. In Section G.5, we describe pilot data that provides an initial validation of both the feasibility of our approach and the practicality of our modeling approach. Finally, in Section G.6, we highlight a few potential applications of the genetic applet technology.

G.2 Theoretical Considerations

The design of the genetic toggle switch, the adjustable-threshold switch and the two-state oscillator are primarily based on calculations made from deterministic ordinary differential equation models of biochemical reactions. The formulation of these equations from the biochemical reactions is a two-step process, each step carrying a specific set of assumptions. In the first step, the individual biochemical reactions are formulated into a set of rate equations as illustrated by the simple example in Fig. 1. The rate equation formulation applies under certain physical conditions. These conditions, which include homogeneity of the reaction mixture and thermal equilbrium of the reacting species [13], will be termed the *macroscopic* formulation. The second step in the modeling is a simplification of the rate equations by application of conservation laws and separation of time-scales—fast reactions are assumed always to be in equilibrium. Thus, the fast variables can be solved immediately in terms of the slow variables.[1] This simplification is illustrated in Fig. 2.

Although the rate equation formulation of a biochemical process is often used to describe biochemical processes [10, 16–21], including gene expression [10, 16, 21], it is not necessarily an accurate description of biochemical events within a single cell. When the system size becomes small or the reacting species become sparsely populated, as occurs in an individual cell, some of the rate equation conditions no longer apply. In reality, the concentration of a gene product within a cell fluctuates in time about a mean value [22,23]. These fluctuations arise from *internal* noise—i.e.,

[1] More formally, the existence of distinct time scales in the reactions, allows the equations to be expanded in a perturbations series [14]. The simplified equations, which arise by preserving only the lowest order terms in the series, describe the system behavior for long time scales [14, 15].

$$
\begin{array}{lcl}
R + D & \underset{k_-}{\overset{k_+}{\rightleftharpoons}} & C \\
C & \overset{k_c}{\rightarrow} & R + D + M
\end{array}
\quad \Rightarrow \quad
\begin{aligned}
\dot{r} &= -k_+ dr + k_- c \\
\dot{d} &= -k_+ dr + (k_- + k_c)c \\
\dot{c} &= k_+ dr - (k_- + k_c)c \\
\dot{m} &= k_c c
\end{aligned}
$$

Figure 1: Formulation of rate equations for a simplified prokaryotic transcription reaction. R, D, C and M represent the RNA Polymerase, the DNA promoter region, the RNAP-promoter complex and the completed mRNA transcript, respectively. Lowercase letters denote the concentrations of each species. In this scheme, the binding of the sigma factor is not included. The steps of open-complex formation, transcription initiation, and polymerization are included in one catalytic step represented by k_c.

the probabilistic interactions of the reacting molecules [13].[2] As explained below, the macroscopic rate equations describe the dynamics of the mean concentration of gene products in a single cell; or, in the case of a colony of cell clones, the rate equations describe the macroscopically observed concentrations of gene products. However, single-cell gene expression is more appropriately modeled as a stochastic process (the *microscopic* formulation). From stochastic modeling, we can obtain measures of the variability in the concentrations of gene products within a single cell.

In this research, we will examine the effects of stochastic fluctuations on the behavior of the genetic networks described below. The issues pertinent to this research are: (i) whether the intrinsic noise in the gene networks will disrupt the predicted behavior of the gene networks; (ii) whether the system can be adjusted to minimize the negative effects of intrinsic noise; and (iii) whether external or internal noise can be utilized to enhance or expand the functionality of the gene networks.

G.2.1 Modeling Noise in Genetic Networks

A typical approach to stochastic modeling of physical systems is to add an *ad hoc* additive or multiplicative noise term to the macroscopic equations. This approach, called the Langevin approach, has been successfully applied to many physical, chemical and engineering systems, both linear and non-linear [13, 24]. However, it has been shown to accurately model stochastic fluctuations only in the case of *external* noise [13]. If the noise is internal, the physical interpretation of the noise term is unclear and can produce incorrect results [13].

Alternately, one may numerically simulate the complete biochemical reactions [25]. In such simulations, the interactions of individual molecules are computed using the rate constants to describe the probability of a particular reaction. Although such an approach may provide accurate statistical data for a particular set of rate constants, it cannot provide analytical relationships between the system parameters and the concentration fluctuations. Since quantitative data concerning the rate constants in a set of biochemical reactions are often sparse, an analytical description of the stochastic fluctuations is especially important. Such a description provides a deeper understanding of the effects of parameter changes on the system behavior.

[2] In contrast, *external* noise refers to fluctuations arising from the application of a random force that is external to the system [13].

$$\begin{aligned} \dot{r} &= -k_+ dr + k_- c \\ \dot{d} &= -k_+ dr + (k_- + k_c)c \\ \dot{c} &= k_+ dr - (k_- + k_c)c \\ \dot{m} &= k_c c \end{aligned} \qquad \Rightarrow \qquad \dot{m} = \frac{k_c d_0 r}{r + (k_- + k_c)/k_+}$$

Figure 2: Simplification of rate equations for a simplified prokaryotic transcription reaction. The total concentration of promoter DNA, d_0, is constant. The resulting equation is analogous to the familiar Michaelis-Menton formulation of an enzymatic reaction.

Here, we take a more analytical approach that applies in general to systems with internal noise. To begin, we formulate the *master equation*, or *M-equation*, directly from the biochemical reactions [13]:

$$\dot{P}(\mathbf{n}, t) = f(P, \mathbf{n}, \mathbf{k}_+, \mathbf{k}_-, \mathbf{S}, \mathbf{R}, \Omega), \tag{1}$$

where $\mathbf{n}$ represents the concentrations of the reacting species, $\mathbf{k}_+$ and $\mathbf{k}_-$ are the macroscopic rate constants for the forward and reverse reactions, $\mathbf{S}$ and $\mathbf{R}$ are the stochiometric coefficients for the forward and reverse reactions, and Ω is the volume of the system. The M-equation describes the time evolution of the *state-transition probability*, $P(\mathbf{n}, t)$. The state-transition probability can be interpreted as the probability that the system will contain concentrations $\mathbf{n}$ at time t given that it contained initial concentrations $\mathbf{n_0}$ at time t_0. Thus, the solution of the M-equation will provide a complete description of the stochastic behavior of the biochemical system. To completely determine the M-equation, only the system volume, the macroscopic rate constants, the stochiometric coefficients and the initial states of the reacting species must be provided.

G.2.2 Solving the M-Equation

For very simple reactions, the M-equation is linear. In this case, an exact solution for the transition probability, providing a complete description of the stochastic biochemical process, can be readily found. However, the M-equation describing most reactions is nonlinear and an exact solution is not possible. Fortunately, a general technique for obtaining the mean, variance, co-variance, auto-correlation function and higher moments of the solution has been developed. This technique, known as the Ω-expansion, requires only that the biochemical system evolves to a single stable fixed point [13].[3] In this method, the M-equation is expanded in an infinite series about a small system parameter. The parameter typically chosen for this expansion is $\Omega^{-1/2}$ (Ω = the volume of the system). Elimination of terms in the expansion of order Ω^{-1} and higher results in a linear Fokker-Plank equation for the state-transition probabilities, the solution of which is a Gaussian stochastic process. Thus, analytical expressions for the mean, variance, co-variance and autocorrelation

[3] The M-equation can be applied to systems with multiple fixed points, but its results are applicable only to a local region around each stable fixed point. Thus, a different approach is required to analyze the global effects of internal noise in a bistable system. This approach is outlined in Section G.3.

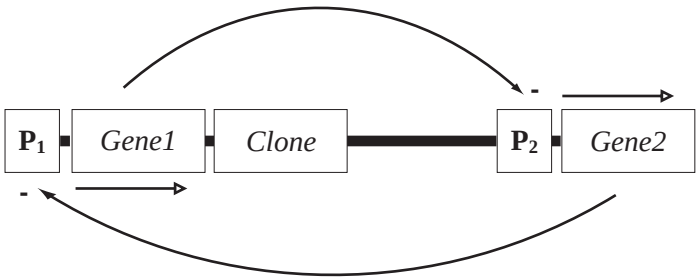

Figure 3: Schematic of toggle switch. Promoter 1, P_1, efficiently transcribes gene 1 unless inhibited by the repressor protein encoded by gene 2. Promoter 2, P_2, efficiently transcribes gene 2 unless inhibited by the repressor protein encoded by gene 1. Open arrows indicate direction of transcription. Clone: an additional gene or genes can be placed under the control of P_1 or P_2.

of the solution—in our case, the concentrations of gene products—are obtained from the original nonlinear master equation. Higher moments can also be found from the Ω-expansion by inclusion of higher orders of $\Omega^{-1/2}$.

One of the more significant results, for this project, of the Ω-expansion concerns the evolution of the mean concentrations of reacting species. In a biochemical reaction described by the M-equation, the Ω-expansion proves that the dynamics of the mean concentration of a species is accurately approximated, to order $\Omega^{-1/2}$, by the macroscopic rate equations [13]. Thus, the Ω-expansion validates the macroscopic modeling approach as an initial approximation of the behavior of a genetic network (or any biochemical system).

G.3 Modeling Genetic Applets

G.3.1 The Basic Toggle Switch

The design for the genetic toggle (see Section G.4) is based principally on the following deterministic model that describes the dynamic interactions of two mutually inhibitory genes. A schematic of the concept is shown in Fig. 3. If properly designed, this system exhibits two stable states. In each state, only one of the two genes is expressed by the host cell. The toggle equations, derived according to the procedures outlined in Section G.2, are:

$$\begin{aligned} \frac{du}{dt} &= \frac{k_1\,\lambda_1/\delta_1}{1+K_{mu}(1+v^{\gamma}/K_{iv}^{\gamma})} - d_1 u \\ \frac{dv}{dt} &= \frac{k_2\,\lambda_2/\delta_2}{1+K_{mv}(1+u^{\beta}/K_{iu}^{\beta})} - d_1 v \end{aligned} \tag{2}$$

where,

u = concentration of gene product 1,
v = concentration of gene product 2,
λ_1 = maximum rate of synthesis of gene 1 mRNA by RNA polymerase,
λ_2 = maximum rate of synthesis of gene 2 mRNA by RNA polymerase,
δ_1 = rate of degradation of gene 1 mRNA,
δ_2 = rate of degradation of gene 2 mRNA,

k_1 = rate of synthesis of gene product 1 by the ribosome,
k_2 = rate of synthesis of gene product 2 by the ribosome,
K_{mu} = Michaelis constant for RNAP binding and transcription of gene 1,
K_{mv} = Michaelis constant for RNAP binding and transcription of gene 2,
K_{iu} = equilibrium constant for inhibitory binding of gene product 1 to promoter 2,
K_{iv} = equilibrium constant for inhibitory binding of gene product 2 to promoter 1,
d_1 = rate of degradation of gene products 1 and 2,
β = degree of multimerization of gene product 1,
γ = degree of multimeriztion of gene product 2.

The first term in each equation describes the synthesis of nascent proteins. Both transcription by the RNA polymerase and translation by the ribosome are included in the first term. Transcription, modeled as in Figs. 1 and 2, is competitively inhibited by the opposing gene product (u or v). Inhibition is achieved by the binding, as a homo-multimer, of one gene product to the opposing gene's promoter region. The multimeric interaction is accounted for by the exponents β and γ in the first term of each equation. The second term describes the rate of degradation of proteins. In the model presented here, this rate is assumed to be similar for both transcripts. Thus, a single rate constant is used for protein degradation.[4] Additional assumptions, implicit in this model, are (i) mRNA turnover is rapid,[5] and (ii) translation of each mRNA transcript occurs at its maximum rate, i.e., proteins are rapidly synthesized from the mRNA by an excess of ribosomes. These assumptions are supported by experimental studies of transcription and translation in prokaryotes [26, 27, 31], and can be extended, with some modifications, to eukaryotes.

Figure 4, which shows the geometric structure of Eqs. 2, reveals the origin of the bistability: the nullclines intersect in three places producing one unstable and two stable fixed points. From this figure, three key features of the system become apparent. First, the nullclines intersect three times, rather than once, because of their sigmoidal shape. The sigmoidal shape arises for β, $\gamma > 1$. Thus, the bistability of the system depends on the homo-multimeric binding of the inhibitory proteins to the DNA. Second, the strengths of the promoters must be matched. If the strengths are not matched, the nullclines will intersect only once, producing a single stable fixed point. Third, the state of the toggle is switched by the application of a transient pulse of an inducing stimulus that pushes the system away from the stable steady state, over the separatrix, and into the opposite basin of attraction.

To build a working genetic toggle switch, it is necessary to understand the effect of each of the parameters in Eqs. 2 on the system behavior. Then, the switch can be designed to produce robust bistable behavior *in vivo*. With eleven parameters, this analysis could be daunting, but by rescaling time and non-dimensionalizing the variables, these equations can be simplified to the following system:

[4] Here we assume identical rate constants for protein degradation to facilitate the description of the toggle behavior. However, this assumption is not necessary. The bi-stable behavior will exist in the toggle under the same conditions on the model parameters, α_1, α_2, β and γ. However, it could alter the balance of the two promoter strengths. Thus, a compensating adjustment in another parameter may be necessary.

[5] If mRNA turnover is not rapid, it will not alter the bistable steady-state behavior predicted by the macroscopic equations. It may, however, affect the kinetics of the system and, hence, alter the effects of internal noise on the fluctuations in expression. Thus, the bi-stability may be altered indirectly.

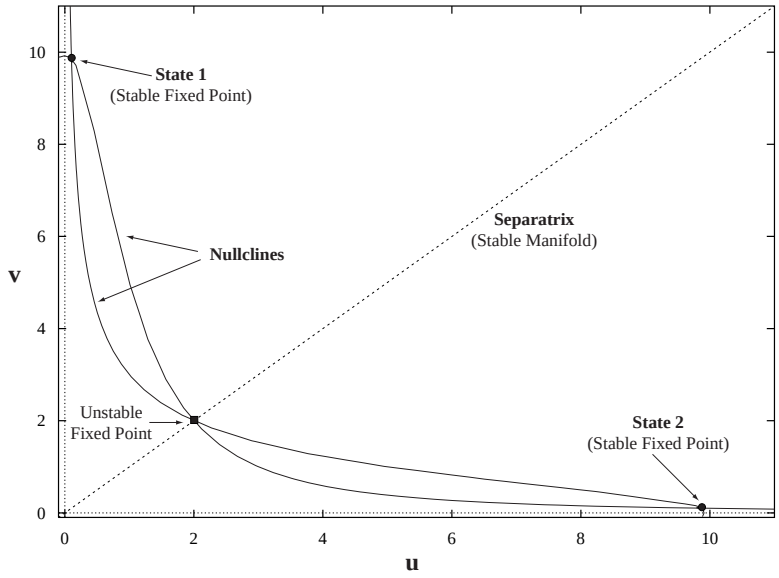

Figure 4: Phase plane diagram of Eqs. 2. A cell with the toggle switch genes will settle to state 1 if its initial state is anywhere above the separatrix; it will settle to state 2 if its initial state is anywhere below the separatrix.

$$\begin{aligned} \frac{d\hat{u}}{d\tau} &= \frac{\alpha_1}{1+\hat{v}^{\beta}} - \hat{u} \\ \frac{d\hat{v}}{d\tau} &= \frac{\alpha_2}{1+\hat{u}^{\gamma}} - \hat{v} \qquad \text{where,} \end{aligned} \tag{3}$$

$$\begin{aligned} \tau &= d_1\, t, \\ \hat{u} &= \frac{u}{K_{iu}\,(1/K_{mv}+1)^{1/\beta}}, \\ \hat{v} &= \frac{v}{K_{iv}\,(1/K_{mu}+1)^{1/\gamma}}, \\ \alpha_1 &= \frac{k_1\,\lambda_1/\delta_1}{d_1\,K_{iu}\,(1+K_{mu})(1/K_{mv}+1)^{1/\beta}} \qquad \text{and,} \\ \alpha_2 &= \frac{k_2\,\lambda_2/\delta_2}{d_1\,K_{iv}\,(1+K_{mv})(1/K_{mu}+1)^{1/\gamma}}. \end{aligned}$$

Nine parameters in the original equations collapse into two. Thus, the range of dynamic behaviors that can be produced by this system is easily understood by analysis of only four parameters. The two new parameters, α_1 and α_2, are defined as the effective strength of promoters 1 and 2, respectively. The effective promoter strength (herein referred to just as promoter strength) is the net effect of the RNA polymerase (RNAP) binding affinity, the rates of transcription and translation, the inhibitor binding affinities, and the rates of degradation of the mRNAs and proteins. These physical quantities can be manipulated in the experimental system to achieve the desired promoter strength (see Section G.4).

Fig. 5 shows the result of two-parameter bifurcation analyses of the system. It is clearly seen in Fig. 5a that the region of bi-stability grows larger as the strength

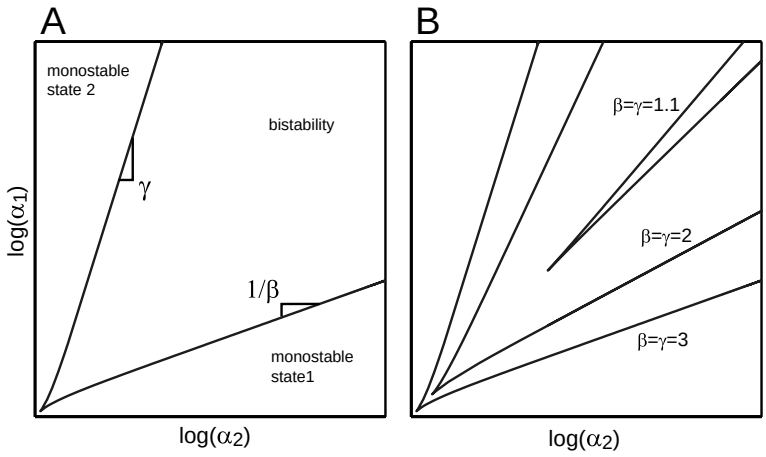

Figure 5: Bifurcation analysis of the toggle. In both plots, the lines separating the bistable region from the monostable regions are saddle-node bifurcations. (A) The slopes of the bifurcation lines, for large α_1 and α_2, are determined by β and γ. (B) As β and γ are reduced, the size of the bistable region decreases.

of both promoters is increased. Though not apparent in the figure, the separation of the two stable states increases. Thus, the system becomes more robust both to imbalances in the promoter strengths and to internal noise. Furthermore, if one promoter is too weak or two strong, then the system falls outside the bistable region in Fig. 5. In Fig. 5b, the bifurcation analysis reveals that, the slopes of the bifurcation lines, for large α_1 and α_2, are determined by β and γ. Thus, to obtain bistability, at least one of the inhibitors must repress expression with cooperativity greater than one. This suggests that repressor multimerization, or multiple operator sites in the promoter, is necessary to obtain bistability. Higher-order multimerization will increase the robustness of the system, allowing weaker promoters to achieve bi-stability.

It should be noted that the above theory, though it is described for a system with a competitive DNA binding inhibitor, applies qualitatively to systems with other kinds of inhibition. Inhibition through protein-protein binding, un-competitive, and non-competitive interactions will result in the same qualitative features of bistability.

G.3.2 Analysis of Internal Noise in the Toggle

As explained in Section G.2, the Ω-expansion approach cannot describe the effects of internal noise in a bistable system. This approach fails because the Ω-expansion can only describe the local effects of stochastic fluctuations. To better illustrate why the Ω-expansion fails, and to illustrate the alternate method used to analyze internal noise in a bistable system, we will utilize the following heuristic description of internal noise in a biochemical system.

The dynamics of certain systems can be visualized as a ball rolling on a potential surface. In a monostable system, such as that given in Fig. 1 in Section G.2, the potential surface has a single dip, a potential well, the bottom of which is centered at the stable fixed point (Fig. 6). In a dissipative deterministic system, the ball will ultimately settle to the very bottom of the well. However, in a system with internal

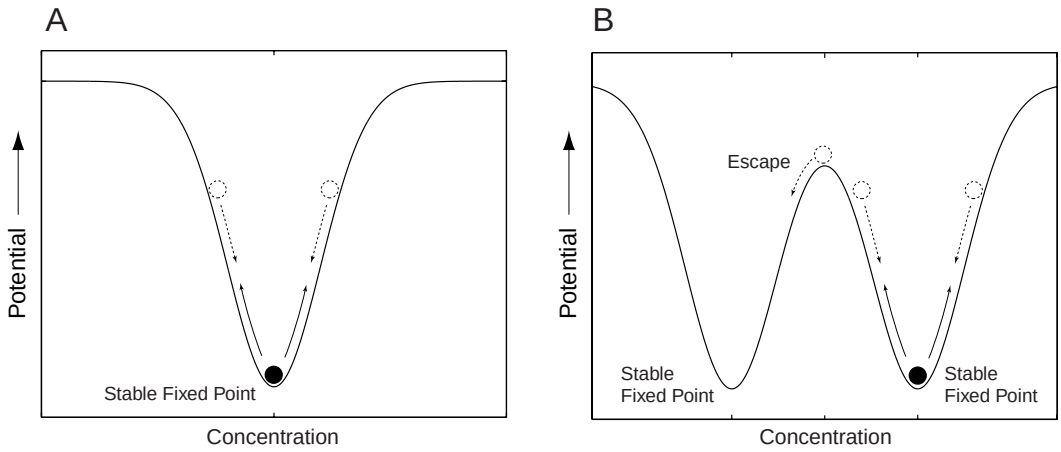

Figure 6: Example potential surfaces. Arrows illustrate the competing effects of the internal noise (solid arrows) and the macroscopic dynamics (dashed arrows). (A) Monostable system. (B) Bistable system.

noise, the ball will never settle because it is constantly kicked away from the bottom. Thus, the system is characterized by a competition between the macroscopic dynamics which pushes the system toward the fixed point and the internal noise which kicks the system away from the fixed point. In a monostable system, internal noise serves only to create the fluctuations around the fixed point. The distance that the ball is kicked away from the fixed point increases as the steepness of the potential well decreases. The fluctuations of the ball about the fixed point are analogous to the fluctuations of a reacting species about its mean concentration. The M-equation and Ω-expansion described in Section G.2 provide a general method for calculating the relationship between the parameters of the biochemical reaction and the shape of the potential well.

In a bistable system, such as the toggle, the internal noise can no longer be viewed as simply creating fluctuations about the mean concentration. Such a system can be visualized as a potential surface with two wells; the bottom of each well corresponds to the two macroscopic stable fixed points (Fig. 6). Thus, the ball will fluctuate around the bottom of one potential well until a sufficiently large kick pushes it over the barrier and into the opposite potential well (where it will remain until another large kick pushes it back). Here, the shape of the well will determine the significance of the internal noise in causing the random jumping between states. However, the system cannot be analyzed with the Ω-expansion approach because the macroscopic equations do not possess a single stable fixed point. The Ω-expansion cannot handle the situation where the ball escapes from a well. Thus, an approach exemplified by *Kramers' escape problem* [13] is necessary. In this approach, the escape time of the ball—the average time for the ball to escape from one potential well—is calculated. The escape time, in terms of the genetic toggle switch, is the average time for a single cell to flip from one expression state to the other due to internal noise. If the escape time from a state in a bistable system is large relative to the time scale of the macroscopic equations, the state can be considered *metastable*. In other words, a bistable system with internal noise will fluctuate about one of the stable states with little or no random switching to the opposing state. If, on the other hand,

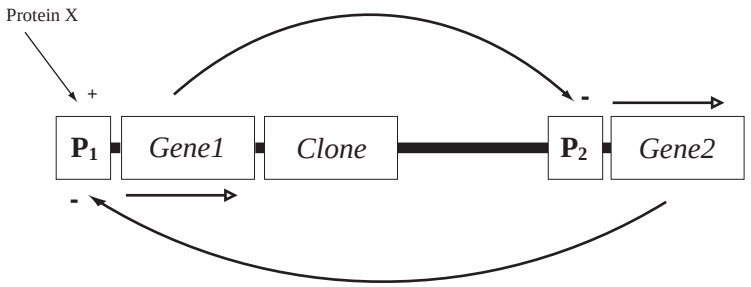

Figure 7: Schematic of adjustable-threshold switch. Transcription of gene 1 by promoter 1, P_1, is activated by protein X and simultaneously inhibited by the repressor protein encoded by gene 2. Promoter 2, P_2, efficiently transcribes gene 2 unless inhibited by the repressor protein encoded by gene 1. Open arrows indicate direction of transcription. Clone: an additional gene or genes can be placed under the control of P_1 or P_2.

the escape time is of the same order as the time scale of the macroscopic equations, the system will be likely to flip randomly between the two stable states. Thus, the bistable behavior predicted by the macroscopic equations will not hold for the real system with internal noise.

In practice, our goal is to design a toggle switch whose bistable behavior is robust with respect to internal noise fluctuations. To this end, we will apply the theories outlined in this section and Section G.2 to analyze the variability in single-cell gene expression. This data, which is obtained as described in Sections G.4 and G.5, will be used to evaluate both the modeling techniques as well as the size and character of the internal noise in the gene networks. These methods will first be applied to simple monostable expression systems, such as those in plasmids pMKN3 and pMKN7 (see Section G.4). These model systems will provide the knowledge and experience necessary to analyze the effects of noise in the more complicated toggle switch. If it is found that the toggle switch behaves poorly as a bistable system due to internal noise, we will refine the toggle as necessary to produce a more robust bistable behavior.

G.3.3 The Adjustable-Threshold Switch

By a relatively minor modification of the toggle design, a genetic switch with an adjustable switching threshold is produced. (Shown schematically in Fig. 7.) Like the toggle, this device is composed of two mutually inhibitory genes. However, promoter 1 is modified such that it cannot transcribe gene 1 without the aid of an additional activator protein. In the absence of protein X, promoter 2 will dominate promoter 1 and gene 2 will be expressed. As the concentration of protein X rises, the strength of promoter 1 will rise as well. Eventually, the strength of promoter 1 will exceed that of promoter 2 and the device will abruptly switch to the expression of gene 1. By manipulating the relative strengths of promoter 1 (when activated) and promoter 2, the activator concentration at which this transition occurs can be altered.

The behavior of the adjustable-threshold switch is illustrated in Fig. 8, which shows the steady-state concentrations of proteins u and v versus the concentration of x for several values of parameters α_1 and α_2. (As for the toggle switch, α_1 and α_2, are the effective strengths of the promoters.) This figure reveals more clearly

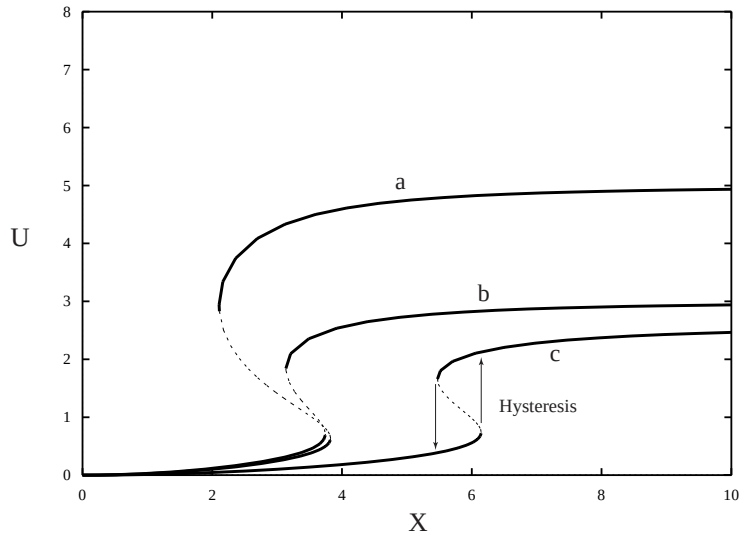

Figure 8: Structure of the threshold. Hysteresis exists at the switching threshold, as indicated by the arrows. Both the location of the threshold and size of the hysteresis can be manipulated independently. Parameter values: $\alpha_1 = 5$, $\alpha_2 = 14$ *(curve a),* $\alpha_1 = 3$, $\alpha_2 = 10$ *(curve b),* $\alpha_1 = 2.7$, $\alpha_2 = 15$ *(curve c).*

the nature of the threshold and the associated hysteresis. It also demonstrates that both the location of the threshold and the size of the hysteresis can be tuned.

G.3.4 The Two-State Genetic Oscillator

A two-state genetic oscillator is a device which alternately expresses one or another of two genes. The oscillations will continue indefinitely and without external stimulus, and the period and amplitude of these oscillations can be tuned.

The oscillator (shown schematically in Fig. 9) is a straightforward extension of the threshold device. Oscillations are produced when feedback with a time delay is added to the adjustable-threshold switch. This can be accomplished by indirectly stimulating the synthesis of protein X (the activator of promoter 1) with a gene product of promoter 2. This scheme creates both the feedback and an adjustable delay.

Fig. 10 illustrates the predicted behavior of the system. By adjusting the time delay, the period of oscillations can be altered. Furthermore, the proper tuning of this time delay and the strength of promoter 3 is critical for the production of oscillations. If the delay is too long or too short, or if the promoter is too weak or too strong, the system will not oscillate; instead, it will settle to a steady-state. Both of these parameters can be manipulated experimentally to achieve the desired behavior.

Though not shown here, oscillations can also be produced from an alternate arrangement of the network in Fig. 9. In this design, promoter 3 is replaced with a constitutively transcribed promoter that is *repressed* by the product of gene 3. Secondly, gene 3 is placed under the transcriptional control of promoter 1.

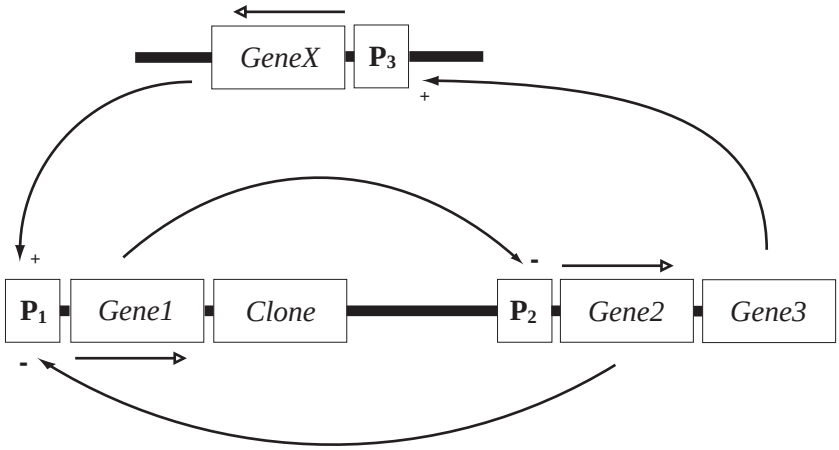

Figure 9: Schematic of two-state oscillator. Promoters 1 and 2 and genes 1 and 2 behave as described in Fig. 7. Transcription of gene x by promoter 3 is activated by gene 3 which is under control of P_2. The activator encoded by gene x stimulates the transcription of gene 1. The time necessary to express gene x creates the time delay needed for oscillations. Open arrows indicate the direction of transcription.

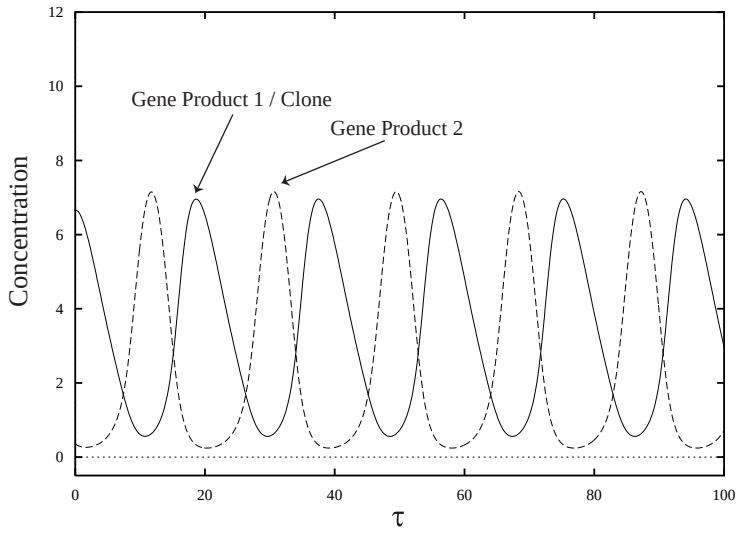

Figure 10: Simulation of the two-state genetic oscillator.

G.4 Implementation

Below, we describe the experimental methods used to construct the toggle switch, adjustable-threshold switch and two-state oscillator. For this project, we choose *E. coli* as the host organism because its gene-regulatory mechanisms are relatively simple and because it facilitates genetic manipulations. Beginning with the toggle switch, all three devices will be constructed and tested. First, we describe the various ways that the devices can be manipulated to achieve the desired behaviors.

G.4.1 Experimental Manipulation of System Parameters

In order to build functioning devices, it will be necessary to experimentally tune various aspects of the gene and protein interactions. For example, the strengths of promoters 1 and 2 in the toggle must be adjusted such that the system falls within the bistable region. Physically, the adjustments can most easily be made to all three systems by manipulating one or more of the following: the strength of RNAP binding to DNA (K_{mu} or K_{mv}), the maximum rate of mRNA synthesis by RNAP (λ_1 or λ_2), the strength of inhibitor binding to the DNA (K_{iu} or K_{iv}), the strength of activator binding to DNA, the rate of translation of mRNA into functional protein (k_1 or k_2), and the rate of protein degradation, i.e., protein stability (d_1). Although the principles described in the theory section apply equally well to prokaryotic and eukaryotic cells, the specific manipulations necessary to adjust the system parameters differ. In this document, only the manipulations that are specific to prokaryotic cells are described.

RNAP Binding. In prokaryotic cells, recognition of the promoter sequence by RNAP is mediated by helper proteins called sigma factors that bind to two sites in the promoter: the Pribnow box (or -10 region) and the -35 region. Each of these sites has an ideal sequence called a consensus sequence. The strength of binding of sigma factors, and thus the strength of RNAP binding, is determined by how closely these regions match their consensus sequence [27]. Furthermore, modifications of a region upstream of the -35 region, called the UP element, have been shown to dramatically alter the rate of transcription [28, 29]. The UP element, which has also been shown to have a consensus sequence, probably enhances the binding of the RNAP complex. By modifying the sequence of the -10, -35 and UP regions, the strength of RNAP binding and, hence, the promoter strength, can be altered.

Transcription Elongation. Once the RNAP binds to a promoter, it opens the DNA double helix and moves forward, adding ribonucleotides to the mRNA transcript. The rate of transcription is determined partially by the nucleotide content and partially by the secondary structure (if any) of the mRNA. High guanosine and cytosine content will slow the transcription rate [27]. Secondary structures that form in the mRNA behind the transcription complex can interfere with the transcription process [27]. Although the DNA content of the coding region cannot be substantially altered (only silent mutations will alter the mRNA sequence without changing the protein properties), a leader region of mRNA can be inserted upstream of the coding region. This region can be designed to slow the rate of transcription elongation.

Inhibitor/Activator Binding. Special sequences of DNA called operators are often found within or near a promoter. The inhibitor proteins (repressors) block

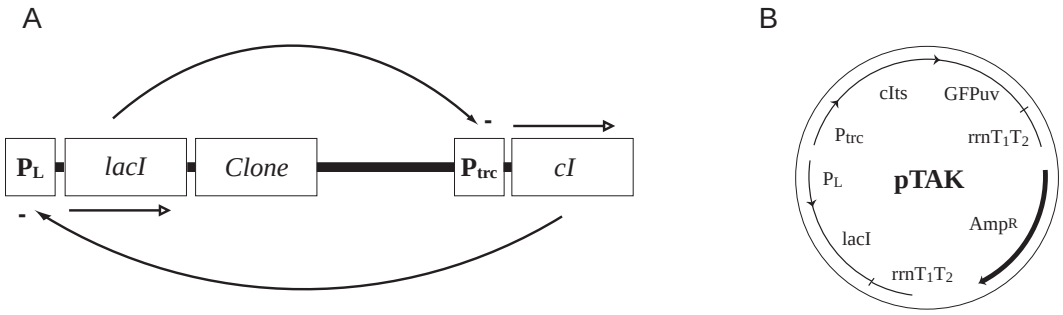

Figure 11: (A) Schematic of toggle switch prototype for expression in E. coli. (B) Physical design of the pTAK plasmid carrying the toggle switch prototype. Key restriction sites are indicated. Promoters and genes and their direction of transcription are indicated. Transcription terminators, $rrnT_1T_2$, are also shown.

transcription by binding to these operators. On the other hand, when an activator binds to an operator, it increases the binding affinity and/or transcription rate of the RNAP. A given repressor or activator will recognize only a specific operator sequence. The affinity of the repressor/activator for the operator can be altered by mutating the operator sequence.

Translation Rate. The rate of translation of mRNA into an amino-acid sequence is governed primarily by three factors: the ribosome binding site (RBS), the secondary structure of the mRNA, and the codon content of the coding region. The RBS is located 5-20 bases upstream of the start codon. Translation is most efficient when this sequence matches a consensus sequence called the Shine-Dalgarno (SD) sequence [27, 30, 31, 33]. Thus, translation rate can be altered by mutating the RBS. As in transcription, the formation of secondary structures by the mRNA can interfere with translation machinery. Thus, modification of the leader region of the mRNA or introduction of silent mutations into the coding region may be used to change translation rate. Finally, in various organisms certain codons are favored, i.e., tRNAs for certain codons are more abundant than others. Translation is more efficient when the favored codons are used [31, 32]. Thus, a coding region can be optimized by introducing silent mutations that utilize the favored codons.

Protein Stability The stability of a protein can be altered by introducing mutations that make it more or less resistant to denaturation or proteolytic degradation. Although it is very difficult to predict a protein's stability, powerful experimental techniques such as directed evolution, DNA shuffling and two-hybrid screening can be used to rapidly screen large numbers of mutant proteins for the desired stability characteristics. Furthermore, recent work has revealed prokaryotic amino acid signal sequences which can be fused to the C-terminal of a protein [34–36]. These sequences target the protein for rapid proteolytic degradation.

G.4.2 Outline of Recombinant DNA Work

The three principal objectives of the initial stages of the experimental work are: (i) to individually test and modify the genetic components of the toggle switch, (ii) to modify and arrange the DNA fragments of the individual toggle components such that they can be integrated into a single plasmid, and (iii) to construct simple genetic expression systems for testing, evaluating and refining the modeling approach.

Repressor	Promoter	Inducer [Co-repressor]
ArsR	Arsenic operon	arsenate or oxidized arsenic, antimony & bismuth
ArgR	Arginine repressor	[L-arginine]
AscG	ASC operon	*unknown*
CscR	Sucrose operon	D-fructose
DeoR (NucR)	Deoxyribose operon	deoxyribose-5-phosphate
DgoR	DGORKAT operon	D-galactonate
FruR	Fructose operon	D-fructose
GalR	Galactose operon	galactose
GatR	Galactitol operon	*unknown*
LexA	SOS response regulon	UV light & RecA protein
RafR	Raffinose operon	raffinose
TetR	Tetracycline resistance operon	tetracycline

Table 1: Alternate *E. Coli* repressors/promoters for the toggle switch.

The components used for the toggle switch prototype are: the P_L promoter and *cIts* gene from the bacteriophage λ, the *lacI* gene from *E. coli*, and the P_{trc} promoter (a fusion of the *trc* and *lac* promoters from *E. coli*). These elements, arranged for the toggle prototype as shown in Fig. 11, are chosen because they are well-characterized and easily obtainable, which facilitates any necessary modifications. Moreover, previous studies have shown that the strengths of the two promoters are reasonably well-matched [37, 38]. The P_L and P_{trc} promoters are constitutive, i.e., they are efficiently transcribed by the RNAP alone, and are repressed by the *cIts* and *lacI* gene products, respectively. The *cIts* gene, because it carries the temperature sensitive mutation, allows switching to P_L expression by thermal induction. If the wild-type allele of the *cI* gene is used instead, then induction of P_L expression can be achieved through the use of nalidixic acid or UV light. On the other hand, switching the toggle to P_{trc} expression is accomplished by a pulse of IPTG. The host organism for the toggle switch, the *E. coli* JM 2.3 strain (*E. coli* Genetic Stock Center strain 5002), is chosen because it is deficient in both a functional *lacI* gene and a lambdoid phage.

As previously mentioned, the choice of the P_L and P_{trc} promoters and the *cIts* and *lacI* genes is not a necessity. Any of the tens or hundreds of trans-acting repressor proteins, and their associated promoters, found in *E. coli* are candidates for use in a genetic toggle switch. A brief search of the Swiss-Prot protein database [39] yielded the several potential repressors listed in Table 1.

Construction of the adjustable-threshold switch and the two-state oscillator will mainly require modifications and additions to the toggle switch. For the adjustable-threshold switch, one of the constitutive toggle promoters will be replaced with a promoter that requires trans-activation. The two-state oscillator requires the addition of a feedback loop to the adjustable-threshold switch consisting of an additional gene and promoter.

Construction of a trans-activated promoter is facilitated by the modular structure of the P_{trc} promoter. The Lac repressor binding site begins exactly at the first nucleotide (+1 nucleotide) of the transcribed DNA. The RNAP consensus sequences

Activator	Promoter	Co-activator
AraC	Arabanose operon	arabanose
CadC	P_{cad} (CAD Operon)	low pH
CRP	*deoP2*	cAMP
CynR	Cyn operon	cyanate
DsdC	Dsd operon	CRP, cAMP
FhlA	Formate dehydrogenase/hydrogenase genes	formate
MalT	*malPp*	maltose
MaoB	Monoamine oxidase gene	CRP, cAMP, tyramine
IlvY	IlvC gene	acetolactate, acetohydroxybutyrate
UreR	Urease operon	urea

Table 2: Potential *E. Coli* activators/promoters for the adjustable threshold switch.

are located upstream of the Lac repressor binding site. Thus, the entire P_{trc} promoter upstream of the +1 nucleotide can be removed and replaced by nearly any trans-activated promoter element. The new hybrid promoter, which retains the Lac repressor binding site, is positively regulated by the trans-activator and negatively regulated by the Lac Repressor.

Many suitable candidates for an activated promoter exist. A quick search of the Swiss-Prot database revealed the possible activators/promoters listed in Table 2. These activator/promoters are all reasonably well-characterized.

Construction of the feedback pathway in the two-state oscillator can be accomplished in one of two ways. The first, as illustrated in Fig. 9, is to place the expression of the activator (protein X) under the indirect control of promoter 2. The single level of indirection, provided by gene 3 and promoter 3, is necessary to produce a required time delay. In the second design (not shown), promoter 3 is replaced with a constitutively transcribed promoter that is repressed by any product of promoter 1.

G.4.3 Quantitative Measurement of Gene Expression

In order to achieve the objectives of this project we require three types of data: macroscopic measurements of gene expression, statistical data describing single-cell gene expression and measurements of the gene-expression dynamics. To this end, we have selected the Green Fluorescent Protein as a reporter of gene expression. The Green Fluorescent Protein, which can form its fluorescent chromophore independently of the host organim, can be quantitatively assayed in *intact* cells. Assays of gene expression are minimally invasive and require few manual manipulations. Thus, quantification of expression using GFP is both efficient and accurate. Macroscopic expression data can be collected using fluorimeters or fluroescence scanners, while single-cell expression data can be collected by flow cytometry or fluorescence microscopy.

Macroscopic Expression Data. In this project, macroscopic expression data is collected using a Molecular Dynamics STORM scanner which excites GFP with a 450 nm blue laser and collects unfiltered emission light using a photomultiplier tube. Bacterial cells expressing GFP are concentrated, transferred to a microtitre

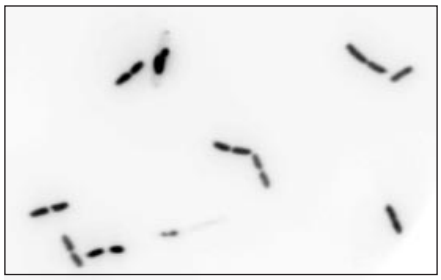

Figure 12: E. coli JM 3.3 stained with Hoechst 33258. Live cells were washed once in PBS (pH 7.4), pelleted, and resuspended in 10 μg/ml Hoechst 33258. Cells were then incubated for 10 minutes in the Hoechst solution, washed once with PBS, pelleted, resuspended in PBS, and transferred to a glass slide. Images were taken using a Zeiss epi-fluorescence microscope with a 100× objective, a 365 nm excitation filter, a 480 nm barrier filter, and a CCD camera.

plate, and scanned. The resulting images are quantified by summing the total fluorescence emission from each well in the microtitre plate. The measured fluorescence is normalized by the cell density, which is measured by the optical scattering at 660 nm (A_{660}). Measurements of expression from the P_{trc} promoter using this method are described in Section G.5.

Single-Cell Expression Data. Measurement of single-cell expression is accomplished using a Zeiss Axioskop epi-fluorescence microscope with a digital CCD imaging system. Single-cell expression is quantified by summing the total expression of each bacterium in the visual field. This task raises two challenges. The first is the object-finding problem, i.e., how to rapidly and consistently distinguish the boundaries of a single bacterium from the noisy background radiation. The second is how to determine the number of copies of the expressing gene in a single bacterium. Since each bacterium may contain a different number of copies of the chromosome (and likewise, the expression plasmid), the expression data will be inaccurate if not corrected for the gene copy number.

We first address the second problem. We assume that plasmid copy number is proportional to the chromosome copy number in a single cell.[6] Thus, we can correct the total single-cell fluorescence by normalizing by chromosome copy number.

The number of chromosomes per cell can be observed under the fluorescence microscope by staining with a DNA specific stain. Figure 12 is an image obtained by staining the bacteria with Hoechst 33258 (Molecular Probes, Eugene, OR). Hoechst 33258 is a membrane permeable stain that binds to the DNA double helix [45–47]. In the image, bacteria with one, two and four chromosomes are visible. Unfortunately, the quantitative measurement of GFP expression simultaneously with a fluorescence stain presents practical difficulties. First, if the emission spectrum of GFP overlaps with the absorption spectrum of the DNA stain, substantial re-absorption of GFP emitted photons can occur. Second, a phenomenon know as Free Resonance Energy Transfer (FRET) can occur if two fluorescent species are in close spatial proximity

[6]Since we are not using stringent plasmids, the plasmid copy number is not necessarily linked to the chromosome copy number [44]. Cell-to-cell variability in plasmid copy number is likely. However, for the reasons outlined in Section G.2, we can assume that the mean copy number *is* proportional to chromosome copy number. The variability in the copy number is accounted for by the stochastic analysis of the expression data. Furthermore, the genes and promoters under study can, if necessary, be inserted into a stringent plasmid.

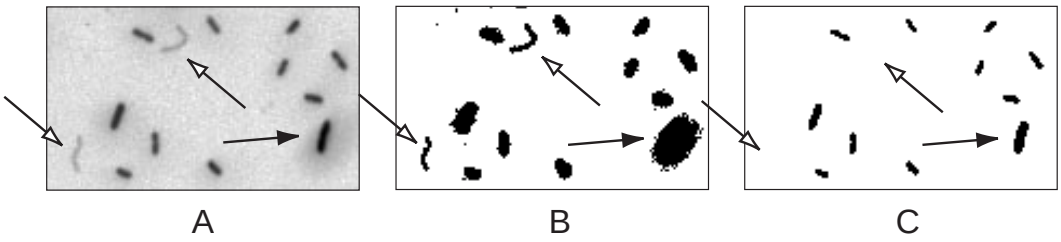

Figure 13: E. coli JM 3.3 expressing GFPuv from plasmid pMKN7. Live cells were washed once in PBS (pH 7.4), pelleted, resuspended in PBS, and transferred to a glass slide. Images were taken using a Zeiss epi-fluorescence microscope with a 100× objective, a 365 nm excitation filter, a 510 nm barrier filter, and a CCD camera. Solid arrows denote an intensely fluorescing cell. Open arrows denote faintly fluorescing cells. (A) Original image—dynamic range is rescaled with a logarithmic transformation. (B) Low-threshold intensity-filter. (C) High-threshold intensity-filter.

and have overlapping emission and absorption spectra [48]. In FRET, the energy absorbed by one species is transferred non-radiatively to the next species which, in turn, radiates the energy fluorescently at a lower wavelength. Either phenomenon, re-absorption or FRET, will alter the emission intensity from GFP and make quantitative measurements difficult.

A way around this problem is found by observing that the length of a rod-shaped bacterium is proportional to the copy number of the chromosome (Fig 12). Thus, we can obtain a measure of chromosome copy number by measuring the length of each bacterium. This measurement is obtained, without DNA staining, by using the image processing algorithm described below.

Image Processing. The image in Fig. 13a shows the GFP fluorecence of individual bacteria. The human eye can easily distinguish the outline bacteria from the moderate background noise. However, it is not clear where to draw the bacterial boundaries so as to obtain a consistent estimate of the total fluorescence. Furthermore, this quantification task must be automated in order to obtain sufficient data for statistical and stochastic analysis.

One method for finding the boundaries would be a straightforward thresholding algorithm in which all pixel intensities below a preset threshold are rejected and all remaining adjacent pixels are considered to be a single bacterium. Although this method is frequently used for object finding, it fails in our case for the reason illustrated in Fig. 13. In the unthresholded image, both faintly (open arrows) and intensely (solid arrows) fluorescing bacteria are visible. If the threshold is set low enough to find the faintest bacteria, it fails to correctly delineate the boundaries of the more intense bacteria. It includes a halo region around the intense bacteria.[7] If, on the other hand, the threshold is raised to eliminate the halos, it also eliminates the faint bacteria.

Fortunately, more sophisticated image processing techniques have been developed that can overcome the problems associated with object-finding in an image with variable background noise. The algorithm we use is illustrated in Fig. 14. The original raw image (Fig. 14a) is first rescaled with a logarithmic transformation to compress the dynamic range (Fig. 14b). Next, the variations in background intensity are removed using an algorithm called unsharp masking [40, 41] (Fig. 14c). Then an

[7] This halo results from optical imperfections, diffraction effects and out-of-focal-plane emissions.

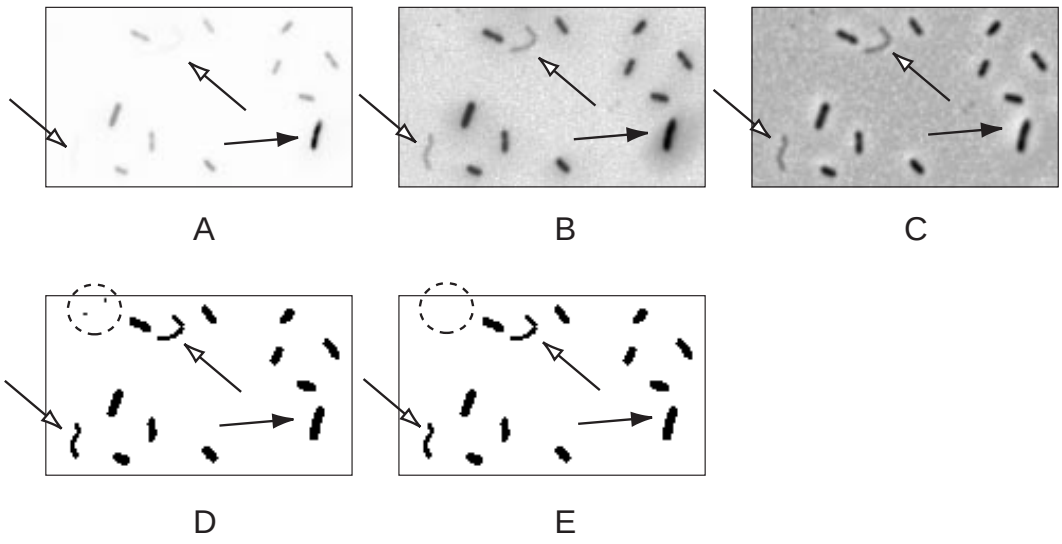

Figure 14: E. coli JM 3.3 expressing GFPuv from plasmid pMKN7. The original image was obtained as in Fig. 13. Solid arrows denote an intensely fluorescing cell. Open arrows denote faintly fluorescing cells. Dashed circles indicate noise objects. All algorithms were implemented in Matlab 5.2 (Mathworks, Natick, MA). (A) Original image. (B) Original image—dynamic range is rescaled with a logarithmic transformation. (C) Image filtered with unsharp masking. (D) Mask produced by edge-preserving smoothing algorithm. (E) Final mask obtained by eliminating noise objects.

edge-preserving smoothing algorithm is applied [42]. This algorithm delineates and intensifies the bacterial boundaries while simultaneously smoothing and rejecting the background noise. Both the faint and intense bacterial objects are highlighted (Fig. 14d). Finally, noise objects that survive the edge-preserving smoothing algorithm are rejected by a minimum size criterion. The resulting mask (Fig. 14e) is then used to quantify the fluorescence and length of the bacterial objects. Single-cell expression data for 662 bacteria obtain through automated image processing of five digital fluorescent micrographs are described in Section G.5.

Dynamic Expression Data. The use of GFP to obtain dynamic expression data suffers from one problem: GFP is an extremely stable protein [32, 34, 43]. The time constant for GFP expression is extremely long—on the order of tens of hours [34]. Thus, as a reporter of gene expression dynamics, it is extremely poor. Other reporters such as cloramphenicol acetyl transferase (CAT) or β-galactosidase (LacZ) or direct assays of mRNA might be used instead. But they do not have the attractive measurement properties of GFP and cannot be used for single-cell assays of gene expression. Fortunately, GFP fusion proteins that have half-lives of less than an hour in *E. coli* have been recently described [34]. These variants consist of the full amino-acid sequence of GFP tagged with a C-terminal protease recognition sequence. Thus, the mutants retain the attractive measurement properties of GFP and have dynamic characteristics that make them suitable for reporting the dynamics of gene-expression.

G.5 Pilot Data

Steady-state gene-expression data collected as described in Section G.4 and the captions of Fig. 15 shows the expression of GFPuv from the P_{trc} promoter of plasmid

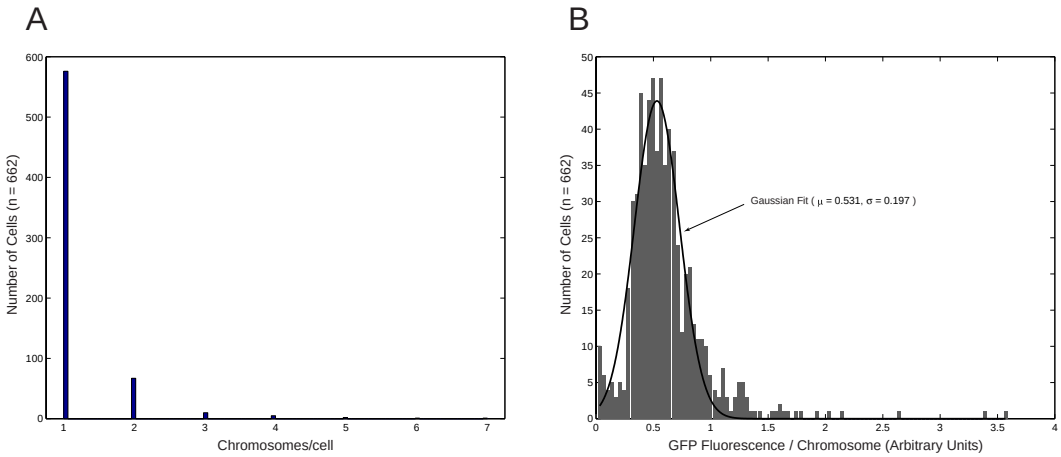

Figure 15: Single-cell GFP expression statistics. Expression of GFPuv from plasmid pMKN7 in E. coli JM 3.3 cells was quantified as described in the text and Fig. 14. 662 cells were quantified from 5 digital micrographs. (A) Histogram of the number of chromosomes per bacterial cell. Number of chromosomes was determined from the cell length. The median cell length was assumed to correspond to a single chromosome. (B) Histogram of single-cell GFP expression illustrating the variability in expression levels.

pMKN7 in 662 *E. coli* JM 3.3 cells. This figure demonstrates the variability in expression among bacteria. The effects of various parameter changes on the variability can be observed and related to theoretical predictions. This type of data will be used to calibrate and refine the theories used to describe the gene networks.

G.6 Applications

Gene Therapy. The treatment of disease through gene therapy will likely require the regulation of transgene expression. Recent work by Rendahl *et al.* [52] demonstrated a successful method for the delivery and controllable expression of a recombinant *epo* gene in mice. Long-term, regulated expression of the *epo* gene, which stimulates the production of red blood cells, may be used as a treatment for hemoglobinopathies or anemia. In this scheme, the *epo* gene is placed under the control of a tetracycline-controlled transcriptional activator. The presence of tetracycline interferes with gene expression by binding to the transcriptional activator. Thus, the expression of *epo* and the consequent production of red blood cells can be turned on and off, by the administration of tetracycline.

The drawback of this approach is that tetracycline must be continuously ingested to maintain the suppressed state of the *epo* gene. Long-term ingestion of tetracycline may be inconvenient or impractical for medical reasons. A better method for the expression of *epo* or any other transgene is to place it under the control of the toggle switch. Expression of the gene will then remain in either the "on" or "off" states until the toggle is switched by the transient ingestion of the appropriate drug.

Another potential gene therapy scheme is to place the transgene under the control of the adjustable-threshold switch. The switch would be constructed such that it activates or inactivates the expression of the transgene in response to changes in the concentration of a particular compound in the body. For example, sufferers of diabetes must inject insulin into their blood-stream when their blood glucose is

abnormally elevated. In a genetic applet treatment of this disease, the liver cells might be transfected with an insulin gene that is under the control of the adjustable-threshold switch. When blood glucose levels are elevated, the switch would initiate production of insulin. Thus, the individual with diabetes would be freed of the need for constant insulin injections.

The concentration of certain hormones in the body fluctuate periodically during the day. Thus, the treatment of diseases in which these hormones are deficient is better accomplished through the periodic ingestion of drugs. Such treatments might be achieved by placing the missing or damaged gene under the control of the two-state oscillator. The period of the oscillator could be adjusted appropriately for that particular gene.

Cell Cycle Control. In recent work, it has been shown that a protein that reversibly binds any one of the cell-division cycle (CDC) proteins can modulate the frequency of cell division or stop and restart cell division completely [53]. This scheme requires the controllable expression *in vivo* of the binding protein. The toggle switch is an ideal system for expressing the binding protein. It can be flipped "on" causing the cell cycle to stop or its frequency to change. The cell will remain in this state until it is desired to restart the cell cycle or return it to its normal frequency. Then, the toggle switch can simply be flipped again. Control of cell division in this manner might be applied to the control of cell growth, used to improve the manufacture of engineered tissues, or possibly used in the treatment of cancer.

The above form of CDC control is passive. That is, the binding protein is a passive element that can only modulate the fundamental dynamics of the cell-division cycle. Though this scheme can speed up the cycle, its stronger effect is to slow or stop the cycle. However, active control of the cell may show even more promise for altering CDC frequency (especially for increasing frequency). In active control, a device with inherent oscillations could be coupled to the cell cycle and drive it at a new frequency. Such a function could be carried out by a two-state oscillator which periodically expresses one of the CDC proteins or a binding protein. Furthermore, the frequency of the two-state oscillator, and hence, the frequency of cell division, could be dynamically controlled by modulating, with external chemical signals, the time-delay in the feedback loop.

Cell Suicide. Once a micro-organism or a human cell is genetically modified with a transgene, it is very difficult to remove that gene. Thus, it may be desirable to remove the altered human cells or destroy the genetically engineered micro-organism. To accomplish this, the genes that initiate apoptosis would be placed under the control of the toggle switch or the threshold switch. When it is desirable to destroy the altered cell, a pulse of the appropriate compound will flip the switch and initiate apoptosis.

A similar scheme has previously been demonstrated by Szafranski *et al.* [54]. In this scheme, streptavidin, which is fatal to the bacterial cell, is expressed only in response to IPTG or depletion of benzioc acid in the cell. However, the threshold at which suicide is initiated cannot be adjusted. In a poorly controlled environment, such as that found outside of the laboratory, trace amounts of the inducing compound may prematurely initiate suicide. Thus, it is desirable to place these suicide genes under the control of the adjustable-threshold switch.

Sensitive Chemical/Protein Sensor and Switch. In the experimental study of gene expression, it may be desirable to monitor, *in vivo* or *in situ*, the concentra-

tion of certain proteins or compounds. Such research could be facilitated by the use of the adjustable-threshold switch. The switch would be designed to express a GFP variant when the concentration of the protein or compound of interest rises above or falls below a particular threshold. A highly sensitive version of this system could be used as the basis for an *in vivo* system for the detection of biological or chemical warfare agents.

The same sort of switch could be used to activate other genes in response to changes in the cellular concentration of a particular protein or compound. For example, the switch could be designed to activate a gene once per cell divison by linking its expression to the concentration of metaphase promoting factor.

G.7 References

[1] Anderson, WF. (1998) *Nature, Suppl.*, **392**, 25–30.

[2] Baron, U, Freundlieb, S, Gossen, M & Bujard, H. (1995) *Nucleic Acids Res.* **23**, 3605–3606.

[3] Bellí, G, Garí, E, Piedrafita, L, Aldea, M & Herrero, E. (1998) *Nucleic Acids Res.* **26**, 942–947.

[4] Forster, K, Helbl, V, Lederer, T, Urlinger, S, Wittenburg, N & Hillen, W. (1999) *Nucleic Acids Res.* **27**, 708–710.

[5] Gossen, M, Freundlieb, S, Bender, G, Müller, G, Hillen, W & Bujard, H. (1995) *Science* **268**, 1766–1769.

[6] Gossen, M & Bujard, H. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 5547–5551.

[7] Lowman, HB & Bina, M. (1990) *Gene* **96**, 133–136.

[8] Poindexter, K & Gayle, RB III. (1991) *Gene* **97**, 125-130.

[9] Schauder, B, Blöcker, H, Frank, R & McCarthy, JEG. (1987) *Gene* **52**, 279–283.

[10] McAdams, HH & Shapiro, L. (1995) *Science* **269**, 650–656.

[11] Arkin, A, Ross, J & McAdams, HH. (1998) *Genetics* **149**, 1633–1648.

[12] Kauffman, SA. (1993) *The Origins of Order: Self-Organization and Selection in Evolution.* (Oxford University Press, New York).

[13] Van Kampen, NG. (1997) *Stochastic Processes in Physics and Chemistry.* (Elsevier, Amsterdam).

[14] Bender, CM & Orzag SA. (1978) *Advanced Mathematical Methods for Scientists and Engineers.* (McGraw-Hill, New York).

[15] Rubinow, SI. (1975) *Introduction to Mathematical Biology* (John Wiley & Sons, New York).

[16] Edelstein-Keshet, L. (1988) *Mathematical Models in Biology* (McGraw-Hill, New York).

[17] Goldbeter, A. (1996) *Biochemical Oscillations and Cellular Rhythms: The Molecular Bases of Periodic and Chaotic Behavior* (Cambridge University Press, Cambridge).

[18] Goldbeter, A. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 9107–9111.

[19] Novak, B & Tyson, JJ. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 9147–9152.

[20] Reich, JG & Sel'kov, EE. (1974) *FEBS Letters* **40**, Suppl., S119–S127.

[21] Smolen, P, Baxter, DA & Byrne, JH. (1998) *Am. J. Physiol.* **274**, C531–C542.

[22] McAdams, HH & Arkin, A. (1997) *Proc. Natl. Acad. Sci. USA* **94**, 814–819.

[23] McAdams, HH & Arkin, A. (1998) *Annu. Rev. Biophys. Biomol. Struct.* **27**, 199-224.

[24] Risken H. (1996) *The Fokker-Plank Equation.* (Springer, Berlin).

[25] Gillespie DT. (1977) *J. Phys. Chem.* **81**, 2340–2361.

[26] Alberts, B, Bray, D, Lewis, J, Raff, M, Roberts, K & Watson, J. (1994) *Molecular Biology of the Cell* (Garland Publishing, Inc., New York).

[27] Darnell, J, Lodish, H & Baltimore, D. (1990) *Molecular Cell Biology* (Scientific American Books, Inc., New York).

[28] Estreem, ST, Gaal, T, Ross, W & Gourse, RL. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 9761–9766.

[29] Yamada, M, Kubo, M, Miyake, T, Sakaguchi, R, Higo, Y & Imanaka, T. (1991) *Gene* **99**, 109–114.

[30] Backman, K & Ptashne, M. (1978) *Cell* **13**, 65–71.

[31] Jacques, N & Dreyfus, M. (1990) *Molecular Microbiology* **4**, 1063–1067.

[32] Clontech Laboratories, Inc. (1999) *Living Colors GFP User Manual.* (http://www.clontech.com/clontech/Manuals/index.html)

[33] Shine, J & Dalgarno, L. (1975) *Nature* **254**: 34–38.

[34] Andersen, JB, Sternberg, C, Poulsen, LK, Bjorn, SP, Givskov, M & Molin, S. (1998) *Appl. Environ. Microbiol.* **64**, 2240–2246.

[35] Keiler, KC, Waller, PRH & Sauer, RT. (1996) *Science* **271**, 990–993.

[36] Keiler, KC & Sauer, RT. (1996) *J. Biol. Chem.* **271**, 2589–2593.

[37] de Boer, HA, Comstock, LJ & Vasser, M. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 21–25.

[38] Rosenberg, M, Yen-sen, H & Shatzman, A. (1983) *Methods in Enzymology* **101**, 123–139.

[39] *Swiss-Prot: Annotated Protein Sequence Database* (http://expasy.hcuge.ch/sprot/sprot-top.html).

[40] Brinkmann, BH, Manduca, A & Robb, RA. (1998) *IEEE Trans. Med. Imag.* **17**, 161–171.

[41] Winter, J (1980) *Radiology* **135**, 234–235.

[42] Shah, J. (1996) In: *Proc. IEEE Conf. on Comp. Vis. and Patt. Recog.*, pp. xvi+932, 136–142.

[43] Tombolini, R, Unge, A, Davey, ME, de Brujin, FJ & Jansson, JK. (1997) *FEMS Microbiol. Ecol.* **22**, 17–28.

[44] Ausbel, FM (1988) *Current Protocols in Molecular Biology* (Wiley Interscience, New York).

[45] Arndt-Jovin, DJ & Jovin, TM. (1989) *Meth. Cell Biol.* **30**, 417–478.

[46] Loontiens, FG, Regenfuss, P, Zechel, A, Dumortier, L & Clegg, RM. (1990) *Biochemistry* **29**, 9029–9039.

[47] Stokke, T & Steen, HB. (1985) *J. Histochem. Cytochem.* **33**, 333–338.

[48] Herman, B (1989) *Meth. Cell. Biol.* **30**, 220–243.

[49] O'Gorman, RB, Rosenberg, JM, Kallai, OB, Dickerson, RE, Itakura, K, Riggs, AD & Matthews, KS. (1980) *J. Biol. Chem.* **255**, 10107–10114.

[50] Dunaway, M, Manly, SP & Matthews, KS. (1980) *Proc. Natl. Acad. Sci. USA* **77**, 7181–7185.

[51] Dunaway, M & Matthews, KS. (1980) *J. Biol. Chem.* **255**, 10120–10127.

[52] Rendahl, KG, Leff, SE, Gillis, RO, Spratt, KS, Bohl, D, Van Roey, M, Donahue, BA, Cohen, LK, Mandel, RJ, Danos, O & Snyder, RO. (1998) *Nature Biotechnology* **16**: 757–761.

[53] Gardner, TS, Dolnik, M & Collins, JJ. (1998) *Proc. Natl. Acad. Sci. USA* **95**, 14190-14195.

[54] Szafranski, P, Mello, CM, Sano, T, Smith, CL, Kaplan, DL & Cantor, CR. (1997) *Proc. Natl. Acad. Sci. USA* **94**: 1059–1063.