

Numerical simulation of stochastic gene circuits

Howard Salis, Yiannis Kaznessis*

Department of Chemical Engineering and Materials Science, Digital Technology Center, University of Minnesota,
421 Washington Avenue, SE, 151 Amundson Hall, Minneapolis, MN 55455-0132, USA

Received 22 September 2003; received in revised form 7 January 2004

Available online 12 October 2004

Abstract

Armed with increasingly fast supercomputers and greater knowledge of the molecular mechanisms of gene expression, it is now practical to numerically simulate complex networks of regulated biological reactions, or gene circuits. Using an *exact* stochastic simulation algorithm, we obtain an accurate time-evolution of the behavior of complex gene circuits, including the effects of fluctuations caused by highly dilute, but significant, regulatory proteins. Specifically, we examine an important gene circuit, the bistable switch, and use the stochastic simulation algorithm to develop design principles that will enable us to produce a fast and robust switch for use in future applications. We pay particular attention to different transcriptional control mechanisms and their effects on the amount of fluctuations in the system, emphasizing methods that increase certainty and improve the switching rate.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Stochastic simulation; Gene circuits; Biochemical networks

1. Introduction

Gene expression is the primary method that a biological organism converts the stored information within its DNA into functional molecules. The dynamic regulation of gene expression is critical to an organism's proper functioning because many of the encoded functional molecules, whether they are structural proteins, enzymes, regulatory proteins, or functional RNAs, have positive effects only when certain environmental stimuli are present or at specific points in the organism's life cycle. The production of unneeded molecules comes at significant cost to the energy reservoir of the organism, in addition to any possible negative effects the molecules may cause when untimely expressed. Regulation may occur at many of the different steps of gene expression, including transcriptional initiation, elongation, and termination (Borukhov & Nudler, 2003; Burgess & Anthony, 2001; Coulombe & Burton, 1999; Dai & Rothman-Denes, 1999; Orphanides & Reinberg, 2002) as well as translational

initiation and termination (Ramakrishnan, 2002; Romby & Springer, 2003). Regulatory proteins or RNA may indirectly or directly alter the binding interactions of specific sequences of DNA or mRNA, either positively or negatively affecting transcriptional and translational efficiency (Franch & Gerdes, 2000; Masse, Majdalani, & Gottesman, 2003; McLeod & Johnson, 2001; Rojo, 2001; Xu & Hoover, 2001). The stabilities of regulatory proteins and mRNAs are also regulated by other proteins, adding an additional layer of complexity (Jenal & Hengge-Aronis, 2003).

Using quantitative modeling, it is now practical to simulate large-scale systems of biological processes, including gene expression, in order to design organisms that possess specific, desired functions. By inserting a system of novel genes into an organism with a desired connectivity, one may create a variety of interesting mini-functions, or motifs.

Some proposed motifs include switches, oscillators, amplitude filters, noise filters or amplifiers, and memory storage (Wolf & Arkin, 2003). Combining these individual motifs, one may create potent applications that have significant potential in both the industrial and medical fields. One significant application is a biosensor, which may be used for the

* Corresponding author. Tel.: +1 612 624 4197; fax: +1 612 626 7246.
E-mail address: yiannis@cems.umn.edu (Y. Kaznessis).

detection of both environmental toxins and disease-indicative chemical species within the human body. The ability to “program” a biological organism to perform specific functions, either with a desired frequency or in response to environmental stimuli, combines the capability of an analog signal processor with the organism’s innate ability to chemically and physically alter itself and its environment. The construction of a designed system of genes that performs a specific function has been termed genetic circuit engineering (Hasty, McMillen, & Collins, 2002) with each motif considered a gene circuit. Currently, many of the proposed designs for gene circuits remain inarticulate, lacking the detail to be directly applied to an experimental system. However, with the use of highly descriptive models and relatively new mathematical formulations, it will be possible to create and design novel gene circuits in silico, guiding their successful construction in an organism.

1.1. Two gene circuits have been successfully constructed in *Escherichia coli*, demonstrating a proof of concept

Two functional gene circuits, the bistable switch and the oscillator, have already been constructed inside biological organisms, demonstrating the ability of multiple genes, through their interconnections, to perform specific tasks (Elowitz & Leibler, 2000; Gardner, Cantor, & Collins, 2000). The quality of the designs, however, may be improved. These gene circuits utilize genetic components from both the lac operon of *E. coli* and the lambda phage virus (Ptashne, 1996). Genetic components may be considered “building blocks” because of the relative ease in which they may be extracted, replicated, altered, and spliced into new biological organisms. As more systems become as well characterized as the lac operon and lambda phage virus, more building blocks will become available for use, making the design of gene circuits easier and more flexible. In addition, through common genetic engineering techniques, already available genetic components may be altered to fit the necessities of a particular design. Novel molecules may also be designed to fit a specific requirement (Ansari, Mapp, Nguyen, Dervan, & Ptashne, 2001).

1.2. The “building blocks” of the lac operon are used to design a bistable switch genetic circuit

The lac operon of *E. coli* has been well studied (Beckwith & Zipser, 1970; Kercher, Lu, & Lewis, 1997; McKnight & Yamamoto, 1992), including quantitative measurements. In this work, the genetic components, or “building blocks”, are extracted from the lac operon with the corresponding kinetic parameters, and are used in order to construct a highly descriptive representational model of gene expression. The main goal of our work is to use the model of gene expression to construct a design for a high-certainty bistable switch genetic circuit, consisting of two genes, each producing a repressor that represses the expression of the other gene (see

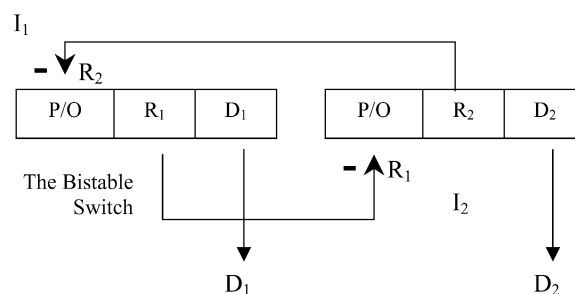


Fig. 1. Gene circuit diagrams of the bistable switch. Each gene module, consisting of a promoter (P), multiple operators (O), and coding gene segments, is represented by a large number of mechanistic reactions. Arrows represent the expression of the gene. Interactions between genes may be activating (+) or repressive (–), through the mechanism of activators or repressors binding to key DNA sites. The i th gene produces a repressor, R_i , and one co-expressed protein, D_i , as the output signal. Inducer molecules, I_1 and I_2 , bind to repressors, R_2 and R_1 , respectively, disrupting their ability to repress gene expression.

Fig. 1). The addition of an inducer, a chemical species that disrupts the repression of a gene, toggles the switch from one state to another.

The major components of the lac operon consist of a promoter sequence, three different operator sites, and three coding genes (see Fig. 2). The Lac repressor, a symmetric tetramer protein, binds specifically to the three operator sites with different affinities and, because of its symmetry, may bind to two operators at once, looping the DNA around itself (Friedman, Fischmann, & Steitz, 1995). Repressors bound to an operator site become an obstacle to successful transcriptional initiation and elongation. Lactose, the inducer for the lac operon, or its more potent imitator isopropyl-thiogalactoside (IPTG) binds to the Lac repressor and disrupts its ability to bind to the operators in a process called induction. Induction results in a rapid increase in gene expression.

1.3. Accurate quantitative modeling of biological systems requires new methods based on stochastic process theory

Biological systems are unique because they feature extremely small length scales, highly dilute chemical concentrations, and event driven processes. A single cell may contain only 10 molecules of a chemical species, making reactions that affect it probabilistic rather than deterministic. When such small numbers of chemical species interact, through reactions, we may no longer assume that the rates of the reactions are continuous, deterministic functions. We must represent the system as a Jump Markov stochastic process, where each reaction represents the transition of the system from one state to another, one where the reactant molecules are consumed and the product molecules are synthesized (for an introduction, see Gillespie, 1992). Reactions will occur discontinuously in time, with their probability of occurrence a function of the number of chemical molecules in the system. The stochastic simulation algorithm, an exact method of

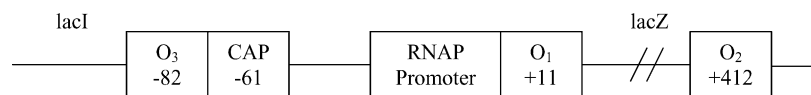


Fig. 2. The transcriptional control components of the *lac* operon of *E. coli*. Locations of the primary operator, O_1 , and auxiliary operators, O_2 and O_3 , are shown relative to the catabolite activating protein (CAP) and RNA Polymerase (RNAP) binding sites. The coding segments for the *lacI* and *lacZ* coding genes are also shown for reference. Numbers refer to the base pair position relative to the transcriptional initiation site within the promoter.

simulating a Jump Markov process, was developed by Gillespie (1976) and has been successfully used to represent the dynamics of gene expression (Arkin, Ross, & McAdams, 1998; Wolf & Arkin, 2002).

In the following sections, we define a genetic circuit bistable switch, list the reactions we use to model its gene expression and regulatory interactions, describe the kinetic parameters used, and explore the effects of changes in design on the performance output of the switch. These results lead us to develop a set of design rules for genetically engineering any high-certainty, fast and robust bistable genetic switch.

2. Model of a bistable switch gene circuit and computational methodology

2.1. Definitions of connectivity and input/output signals

The bistable switch gene circuit consists of two genes, each producing a repressor that represses the expression of the other gene. Each gene contains one additional coding segment that produces a protein, which may be an enzyme or fluorescent protein. One defines the switch as “On” when the number of molecules of repressor two is greater than that of repressor one and “Off” when the reverse is true. The switch gene circuit allows one to program a decision-making branch point in a biological system. Switching between the on and off states requires the presence of inducers. The “Off” inducer disrupts the “On” repressor’s ability to repress the “Off” gene, resulting in the switch changing its state to “Off”. The “On” inducer similarly changes the switch’s state by disrupting the “Off” repressor. The input signal is the number of molecules of each inducer. The resulting output signal consists of the number of “Off” or “On” repressors or co-expressed proteins. The co-expressed proteins may be selected in order to specifically interact with a downstream system, becoming input signals to another gene circuit, or to provide a detectable response, such as fluorescence. Using a switch gene circuit, one can create a decision-making branch point. If the “On” inducer is present, the switch toggles from producing the “Off” co-expressed proteins to producing the “On” co-expressed proteins. If the “Off” inducer is present, the opposite occurs. Once the transition occurs, an ideal switch will maintain its state without the presence of additional inducer. This is entirely analogous to an If-Then-Else construct in computer programming.

However, owing to the significant fluctuations in biological systems, the bistable switch must be carefully designed to produce a usable output signal. Transcriptional control is

not simply a hard switch and is sensitive to the design of the promoter and operator sequences. In order to create the best switch, the transcriptional control mechanism must provide tight enough repression. The lack of tight repression is called “leaky expression” and ruins the output signal of the bistable switch.

2.2. A detailed list of reactions that forms the model of gene expression and regulation

The representational model of a single gene in this work consists of 40 individual elementary reactions and 27 unique chemical species, forming a mechanistic model of the biological machinery of bacterial gene expression, the protein interactions of the repressor, and the mechanisms of repression. To create a gene circuit, one determines the connectivity between each gene, including any molecular interactions between transcriptional or translational machinery and gene products. Reversible reactions are represented by two opposing irreversible reactions, because transitions from state to state in a Markov model must be unidirectional.

Our model of bacterial gene expression is similar to previous stochastic simulations (Arkin et al., 1998), except for a few notable exceptions. Transcription initiation is represented in a fully non-equilibrium manner, where the binding and unbinding of the RNA Polymerase and repressors to the DNA are individually represented. The alternative approach of assuming equilibrium may not always be valid for highly dynamic systems.

Four reactions describe the dimerization and tetramerization of the repressor monomers. A first order reaction describes the degradation of the repressor. Six elementary reactions describe transcriptional initiation, including RNA Polymerase binding to the promoter region, the closed-to-open conformational change, and RNA Polymerase forward movement across free operators. Transcriptional and translational elongations are represented as many first order reactions in series, each reaction representing the movement of the RNA Polymerase or ribosome across a single nucleotide or codon.¹ Abortive initiation, a process where the RNA Polymerase-DNA complex becomes disassociated when stalled by a repressor-operator complex, is described

¹ Whereas a single reaction is stochastically represented with an exponentially distributed probability of occurrence, multiple first order reactions may be equivalently represented with a single gamma distributed probability of occurrence. Without this equivalence, representing transcriptional elongation would require as many reactions as nucleotides transcribed (Gibson & Bruck, 2000).

by four first order reactions, one for each operator site. Translation initiation is governed by four reactions, describing the competitive binding between ribosome and RNase E for the 5' region of the mRNA. If a ribosome binds to the mRNA, translational elongation is initiated. If an RNase E binds to the mRNA, the mRNA may no longer be bound by additional ribosomes and becomes degraded.

There are three operators adjacent to the promoter region, each capable of binding a repressor produced by the other gene. There is a fourth operator at the end, capable of binding a repressor produced by its own gene. The fourth operator is responsible for self-repression, limiting the expression of any one gene to a maximum rate. In order for the RNA Polymerase to melt into the DNA and initiate transcription, the first operator must be free. If repressors occupy the second, third, or fourth operators, the initiated RNA Polymerase may not move forward. The longer the RNA Polymerase remains stalled in the pre-elongation mode, the greater the probability that dissociation occurs and initiation is aborted. A repressor may bind to each individual operator and up to two operators at once, causing DNA looping. There are eight reactions describing repressors binding to individual operators and 12 reactions describing all possible pairs of repressor–operator complexes that may form DNA loops. The fourth operator may not engage in DNA looping.

Each gene produces a repressor and one co-expressed protein. Once the RNA Polymerase has passed the fourth operator, there is no further transcriptional elongation control. There is no additional translational control as well. The co-expressed protein is the output signal for each gene, determining the level of expression.

2.3. Model assumptions

1. The cell is assumed to be a homogenous, well-stirred medium with a large number of water molecules, which regularly collide with the reactant species in the system. In order to use the stochastic simulation algorithm, the velocity distribution of reacting species in the system must return to a Maxwell–Boltzmann distribution after each reaction, requiring them to engage in many non-reactive collisions prior to a reaction occurring (Gillespie, 1976).
2. All housekeeping genes are constitutively expressed and maintained at normal concentrations. This includes all metabolic enzymes not associated with the gene circuit and proteolytic enzymes necessary for protein degradation. Transcriptional and translational elongations are not rate limited by the availability of nucleotides and amino acids. The concentrations of RNA Polymerase and ribosome are large, but, because of the large number of genes being expressed in the cell, the *available* number of RNA Polymerase and ribosome is small and constant. All simulations use 180 and 300 molecules of *available* RNAP and ribosome, respectively.
3. All sigma factors, proteins that globally regulate many genes, are present in constitutive and constant amounts

and do not favor the transcriptional or translational efficiency of any gene in the gene circuit. Sigma factor binding to RNA Polymerase is fast and does not rate limit transcription initiation.

4. Each step of transcription and translation elongation occurs at an average probability, neglecting changes in probability due to differences in nucleotides or codons. There is no transcription or translation elongation regulation, such as Rho-dependent termination or early mRNA hairpin formation.
5. An average of 20 proteins is produced per mRNA transcript. Because of the stochastic competition between the ribosome and RNase E, the resulting standard deviation of production is 19.1. The distribution of the production of proteins is determined by the ratio of ribosomal and RNase E activities.
6. The dimer form of the Lac repressor does not effectively bind to an operator, due to competitive binding from the more active tetramer form (Levandoski et al., 1996).
7. Induction of repressors is complete and instantaneous when inducer is added to the system. The inducer immediately binds all repressors capable of being induced causing an immediate change in binding affinity to all affected operators.
8. The volume of the cell remains constant at 1×10^{-18} L. Because the bistable switch quickly assumes a state, dilution effects from cell growth do not significantly affect the decision-making process.

2.4. Model kinetic parameters

There are two categories of kinetic parameters that have been used. The first describe interactions that current techniques may not alter without destroying the necessary function. From a design perspective, these parameters are constraints in which an engineer may not currently adjust. The second category includes all kinetic parameters that may be adjusted by modifying the interactions of the participating molecules. With these parameters, there is greater flexibility in the design of the gene circuit. For these simulations, the only design parameters are the strength of the repressor–operator affinity, the inclusion of one or more operators, and the relative positioning of each operator within the gene. All other kinetic constants are considered constraints and are kept fixed at their measured values. If no measured value is available from the lac operon, a measurement of the same reaction from a common biological system is used. The parameters are summarized in Table 1.

2.5. Constrained parameters

Transcriptional and translational kinetic parameters are kept constant throughout all simulations. The RNA Polymerase has an affinity of $2 \times 10^7 \text{ M}^{-1}$ with the promoter and a forward binding constant of $2 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$. The irreversible melting process, whereby the RNA Polymerase

Table 1
Kinetic parameters of gene expression machinery and repressor dynamics

Constrained parameters (all genes)		
RNAP binding promoter	$2 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$	McKnight (1992)
RNAP unbinding promoter	0.01 s^{-1}	McKnight (1992)
RNAP transcribing rate	30 nt/s	Vogel (1994)
Stalled RNAP abortive initiation	$5.776 \times 10^{-4} \text{ s}^{-1}$	Common: 10–30 min; $t_{1/2}$
Ribosome binding to mRNA	$0.167 \text{ M}^{-1} \text{ s}^{-1}$	Adjusted to give 20 proteins per mRNA
RNAse binding to mRNA	$0.015 \text{ M}^{-1} \text{ s}^{-1}$	
Ribosome translating rate	100 aa/s	Sorensen (1991)
Repressor degradation	$2.31 \times 10^{-3} \text{ s}^{-1}$	Elowitz (2000)
Co-expressed protein degradation	$3.85 \times 10^{-4} \text{ s}^{-1}$	Typical value
Design parameters		
Repressor monomer dimerization and tetramerization	$k^1 = 10^9 \text{ M}^{-1} \text{ s}^{-1}$ $k^{-1} = 10 \text{ s}^{-1}$	Levandowski (1996)
Repressor binding to operator	$5 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$	McKnight (1992)
Repressor–operator complex dissociation	$2.5 \times 10^{-3} \text{ s}^{-1}/3.85 \times 10^{-4} \text{ s}^{-1}$	Function of operator sequence
DNA loop formation	$5 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$	Function of relative
DNA loop complex dissociation	$1 \times 10^3 \text{ s}^{-1}/1.0 \text{ s}^{-1}$	Distance
Repressor binding to fourth operator	$\text{None}/5 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$	McKnight (1992)
Repressor–fourth operator complex dissociation	$\text{None}/3.85 \times 10^{-3} \text{ s}^{-1}/1.925 \times 10^{-3} \text{ s}^{-1}/7.7 \times 10^{-4} \text{ s}^{-1}$	Function of operator sequence

unwinds the DNA and begins transcription, occurs at a rate of $1 \times 10^{-2} \text{ s}^{-1}$ (McKnight & Yamamoto, 1992). Movement of the RNA Polymerase across unobstructed DNA, including both operators and coding segments, takes place at a rate of 30 nucleotides/s (Vogel & Jensen, 1994). RNA Polymerase molecules that are prevented from moving forward by repressor–operator complexes may undergo abortive initiation with a probability of $5.776 \times 10^{-4} \text{ s}^{-1}$, equivalent to a half-life of 20 min. The coding segment for both the repressor and co-expressed protein is 600 nucleotides long. Ribosome and RNAse E competitively bind to the 5' end of the mRNA at binding constants of 0.167 and $0.015 \text{ M}^{-1} \text{ s}^{-1}$, respectively, producing an average of 20 ± 19.1 proteins per mRNA transcript. Translation elongation occurs at an average rate of 100 codons s^{-1} (Sorensen & Pedersen, 1991). The mRNA transcript has a length of 200 codons, corresponding to 200 amino acids. The repressor and co-expressed protein have half-lives of 5 and 30 min, respectively. The repressor obtains its fast half-life by genetic insertion of a tag that recruits the cell's proteasome machinery (Elowitz & Leibler, 2000).

Lac repressor dimerization and tetramerization equilibrium constants vary with environmental conditions and range from 10^8 to 10^{15} M^{-1} (Levandoski et al., 1996). At these equilibrium constants, the tetramer form is highly favored. In order to design a switch for a less than desirable environmental condition, forward and reverse kinetic constants of $10^9 \text{ M}^{-1} \text{ s}^{-1}$ and 10 s^{-1} , respectively, were used, giving an equilibrium constant at the very low end of the possible spectrum.

2.6. Design parameters

Numerous experiments have altered operator sequence and placement to observe the effects on the

repressor–operator interactions, including DNA loop formation (Müller, Oehler, & Müller-Hill, 1996). The three natural operators in the lac operon, O_1 , O_2 , and O_3 , each have differing affinities to the Lac repressor. In addition, one may use genetic engineering techniques to introduce mutations in the sequence of each operator, fine tuning the affinities to the repressor. The genetic circuit engineer then may select an operator from one of many in order to obtain the desired affinity to the repressor. For these simulations, two types of operator sequences have been used. Both operator sequences bind to the repressor with a forward kinetic constant of $5 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$ (McKnight & Yamamoto, 1992). The dissociation constants are 2.5×10^{-3} and $3.85 \times 10^{-4} \text{ s}^{-1}$, equivalent to macroscopic half-lives of 5 and 30 min, respectively. From a probabilistic perspective, macroscopic half-lives determine the amount of elapsed time when a complex has a 50% chance of decomposition. The genetic circuit engineer may also fine-tune the extent of DNA looping by adjusting the distance between the two operators, altering the probability of loop formation. For the lac operon, the maximum probability of DNA looping occurs at a distance of 70 base pairs. The extent of loop formation decreases with both longer and shorter distances, oscillating with a periodicity of 11 base pairs (Müller et al., 1996). Three different distances between operators were chosen to correspond to three different designs, each employing a varying degree of DNA looping. The first utilizes no DNA looping, while the second and third have equilibrium constants of 5.0 and $5 \times 10^3 \text{ M}^{-1}$, respectively. Forward looping kinetic constants are, for the latter two designs, $5 \times 10^3 \text{ M}^{-1}$.

2.7. Design goals

Two genes are mechanistically represented, as described above; each coding for the production of a repressor and

co-expressed protein of different functional behavior. The repressor produced by the first gene, R_1 , may bind to the first, second, and third operator of the second gene and the fourth operator of the first gene. The repressor produced by the second gene, R_2 , may bind to the first, second, and third operator of the first gene and the fourth operator of the second gene. The first and second genes produce co-expressed proteins named D_1 and D_2 , respectively.

The switch has two significant dimensionless numbers; they are state and certainty. The state of the switch is either “On”, where the number of $D_2 > D_1$, or “Off” otherwise. The certainty of the switch determines to what extent the output signal, either on or off, is reliable and accurate. The certainty of the switch is absolute when the difference in concentration between the co-expressed proteins is highest. The certainty of the switch may be described by $C = |D_2 - D_1| / (D_2 + D_1)$. At an ideal certainty of one, the switch is either fully on or off. At a certainty of zero, the concentrations of both co-expressed proteins are equal and it is impossible to determine the state of the switch. *One goal in designing a better switch is to maximize the amount of certainty in the output signal.*

Another goal in designing a better switch is to decrease the time necessary to switch from one state to another. The amount of time to effectively switch states is inversely related to the numbers of repressor and co-expressed protein at the current state and the rate at which they are degraded. One may either increase the degradation rates of repressor and co-expressed protein or decrease the maximum number of molecules of both products that exist at steady state. The former is possible only if there is a genetic engineering technique to target both the repressor and co-expressed protein for faster degradation. The latter involves only the repressor and the addition of a previously known operator sequence. This latter method was implemented in the design by the inclusion of a fourth operator. One may fine-tune the steady state concentration of repressor and co-expressed protein by modifying the affinity of the fourth operator to its repressor, either increasing or decreasing the magnitude of self-repression desired.

2.8. Computational methodology

Computational simulations were performed using the stochastic simulation algorithm (SSA), created by Gillespie (1976), with substantial enhancements and modifications (Gibson & Bruck, 2000). The algorithm generates a time evolution of a homogeneous system of coupled chemical reactions, explicitly accounting for the fluctuations that are inherent in systems containing few reactant molecules. We have implemented the algorithm using Fortran 95. Analysis of data was carried out using MATLAB.

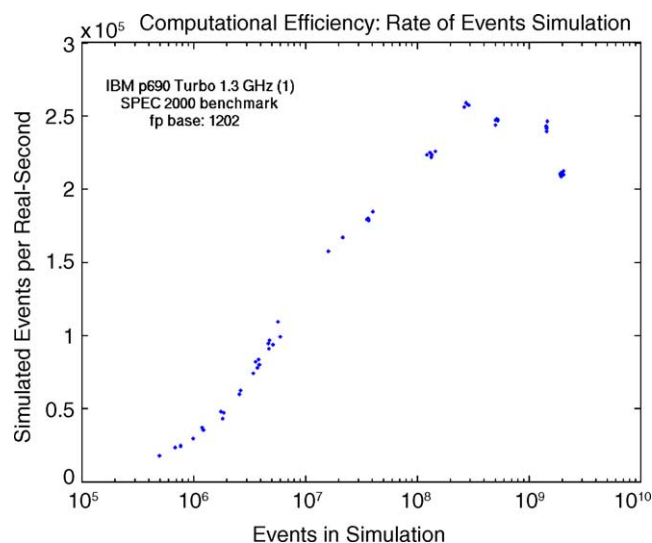


Fig. 3. The rate at which events are simulated is plotted vs. the total number of events in a single trial. The optimizations implemented favor simulations requiring numerous events.

2.8.1. The enhanced stochastic simulation algorithm is capable of efficiently simulating a large-scale system of coupled chemical and biochemical reactions

Simulations were run on a single processor of an IBM pSeries 690 Turbo. Each computational experiment consists of 100–1500 identical trials. The typical trial ranged between 10 and 250 million simulated events. The algorithm is more efficient when simulating large numbers of events. A simulation of 9.8 million events ran at 139,000 events per real-time second while one of 304 million events ran at 223,000 events per real-time second (see Fig. 3).

2.8.2. An overview of the “Next Reaction” variant of the stochastic simulation algorithm

In order to apply the SSA, we must first define the system. The system consists of N unique chemical species and M first or second order reactions in a homogenous volume, V . The state vector X_i , for $i = 1, \dots, N$, contains the numbers of molecules of the i th chemical species. The i th reaction has a propensity, a_i , analogous to a macroscopic rate. Using the reaction propensity, we then produce a joint probability density, $P_i(\tau)$, for each reaction from $i = 1, \dots, M$.

$P_\mu(\tau) d\tau$ is the probability at time t that the μ th reaction has not occurred within $(t, t + \tau)$, and that the μ th reaction in V will occur in the differential time interval $(t + \tau, t + \tau + d\tau)$.

The joint probability density of reaction μ is the product of the two possible events, the latter *conditioned*, or dependent, on the former being true. The first condition, that the reaction may not occur, exponentially decreases with time, and is

$$P_\mu^0(\tau) = \exp(-a_\mu) \quad (1)$$

The second condition, that the reaction will occur within the differential time interval, is equal to the reaction propensity multiplied by the differential element, $a_\mu d\tau$. The product of

the two events produces the joint probability distribution for the μ th reaction.

$$P_{\mu}(\tau) d\tau = a_{\mu} \exp(-a_{\mu} \tau) d\tau \quad (2)$$

We may calculate the reaction propensity using only the macroscopic rate constant, k_i , the state vector \mathbf{X} , and the order of the reaction. Analogous to mass action kinetics, the reaction propensity is

$$a_i = h_i c_i \quad (3)$$

where h_i is the number of *distinct* combinations of reactant molecules for the i th reaction, and c_i is the average specific reaction propensity for the i th reaction.

The order of the reaction affects the calculation of both h_i and c_i in the following manner:

For a first order reaction, $A \rightarrow B$

$$h_i = \#A$$

$$c_i = k_i$$

For a monomolecular second order reaction, $A + A \rightarrow B$

$$h_i = \frac{\#A(\#A - 1)}{2}$$

$$c_i = \frac{2k_i/V}{Na}$$

For a bimolecular second order reaction, $A + B \rightarrow C$

$$h_i = \#A\#B$$

$$c_i = \frac{k_i/V}{Na} \quad (4)$$

where Na is Avogadro's number and $\#X$ refers to the number of molecules of species X .

Using the joint probability density, we can now calculate the *time* at which the next reaction will occur. Once we know both the time and the type of reaction that will occur next, we may simply execute the reaction, change our state, and repeat our calculations again to find the *next* reaction that occurs. To do this, we must first integrate the joint probability density from zero to τ and find the joint probability distribution, $F_{\mu}(\tau)$.

$F_{\mu}(\tau)$ is the probability at time t that the μ th reaction will occur within the interval $(t, t + \tau)$.

$$F_{\mu}(\tau) = 1 - \exp(-a_{\mu} \tau) \quad (5)$$

The joint probability distribution may assume all values between 0 and 1, a property that is shared with a uniform random number of the same domain. We may then use a Monte-Carlo technique to simulate the uniform sampling of probabilities by equating Eq. (5) with a uniform random number (URN) with a domain (0,1). Equivalently, we may also

use $1 - \text{URN}(0,1)$ as our uniform random number. By inverting with respect to τ , we may now obtain the absolute *time* in which the μ th reaction will occur, or

$$t_{\mu} = \frac{-\log(\text{URN}(0, 1))}{a_{\mu}} + t \quad (6)$$

where t is the current simulation time.

In order to find the reaction that will occur next, we must calculate the times at which every reaction occurs and find the shortest time. The reaction with the shortest time is executed; increasing and decreasing the molecules of the product and reactant species, respectively, within the state vector \mathbf{X} . In order to update the time, we may now simply equate the current simulation time with the new "Next" reaction time because of the shift from relative to absolute units. We then repeat the former calculation of reaction propensities and next reaction times, find the shortest next reaction time, and execute that reaction.

Gibson and Bruck (2000) created the "Next Reaction" variant and additional optimizations that improve the efficiency of the stochastic simulation algorithm. Using a dependency graph, the reaction propensities are only recalculated if they have changed. By using an indexed priority queue, finding the minimum of the list of reaction times is optimized. Finally, after recalculating reaction propensities, the reaction times for unexecuted reactions are rescaled rather than requiring newly generated pseudorandom numbers. The latter modification is an application of the random variable transformation (RVT) theorem and reduces the computational intensity involved in the generation of pseudorandom numbers. Combined, these modifications produce a highly efficient algorithm capable of simulating large-scale systems of chemical and biochemical reactions.

3. Results

The following computational experiments analyze the effects of different design choices. In each experiment, numerous trials are performed to obtain an accurate distribution of the possible outcomes and an accurate ensemble average. Each trial begins with the bistable switch in a "null" state, neither on nor off. Because both genes use equivalent kinetic parameters, there is a 50% chance of the bistable switch assuming either state. Of course, small deviations are expected due to the practical limit on the number of trials able to be performed in a reasonable amount of time. The results are depicted using both histograms and time series graphs. The histograms represent the distribution of outcomes over many trials. The height of each bar is equal to the percentage of trials in a particular state, either on or off, at a particular value of certainty. The ideal switch would have two bars, each at 50%, with a certainty of one. Increasing the certainty of a switch will shift the distribution of outcomes towards the right. The time series graphs depict both the ensemble mean and standard deviation of each experiment.

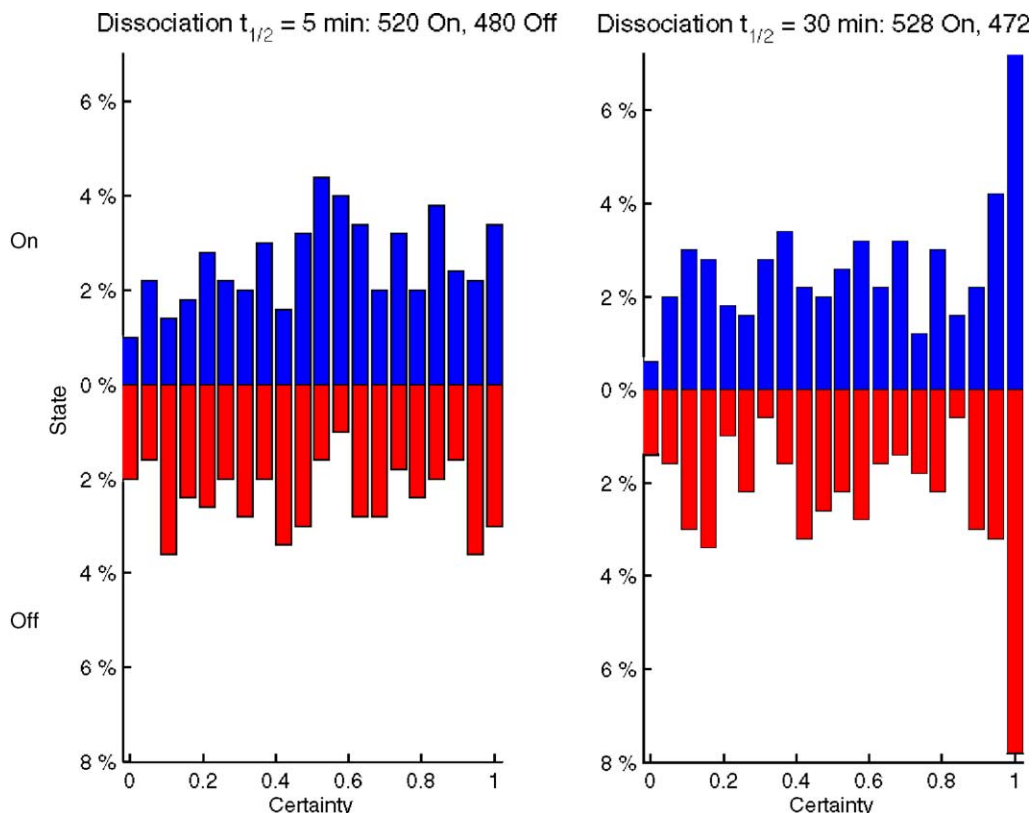


Fig. 4. A state vs. certainty histogram of a one operator design with repressor–operator dissociation constants of 2.5×10^{-3} and $3.85 \times 10^{-4} \text{ s}^{-1}$, equivalent to macroscopic half-lives of 5 and 30 min, respectively.

3.1. The effect of increasing the repressor–operator affinity

Transcriptional control is dependent upon the high affinity of a repressor for an operator. By increasing the affinity, one decreases the probability that a repressor will unbind from an operator and cause “leaky expression”. In a single operator design, the dissociation constant of the operator is decreased from 2.5×10^{-3} to $3.85 \times 10^{-4} \text{ s}^{-1}$, equivalent to macroscopic half-lives of 5 and 30 min, respectively.

With a half-life of only 5 min, the first design features very leaky transcriptional control, resulting in a poor switch (Fig. 4). In this design, only 22.5% of the trials have a certainty of 0.8 or greater, which is only 2.5% greater than a uniformly random distribution. By increasing the affinity of the operator to an equivalent half-life of 30 min, one greatly increases the certainty of the switch, resulting in 29.8% of the trials having a certainty of 0.8 or greater.

3.2. The effect of multiple operators

In natural biological systems, the usage of multiple operators is common. The reason behind the usage of multiple operators may be understood by comparing the certainty of the output signal across different designs. In three separate designs, simulations are run with either one, two, or

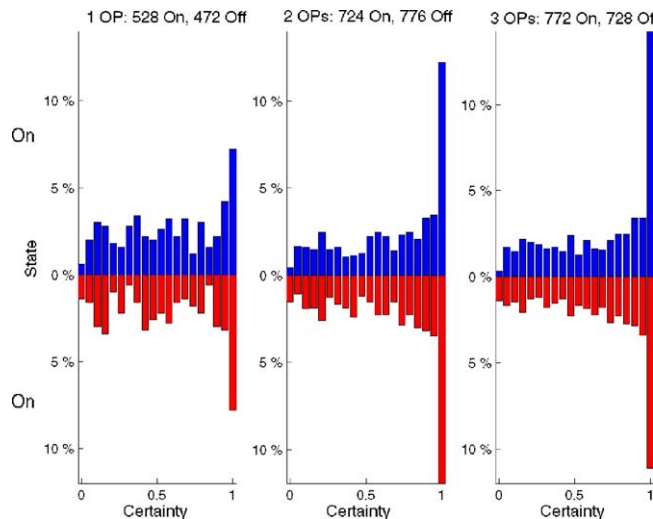


Fig. 5. A state vs. certainty histogram of a one, two, and three operator design, each with repressor–operator dissociation constants of $3.85 \times 10^{-4} \text{ s}^{-1}$.

three working operators on both genes (Fig. 5). All operators have dissociation constants of $3.85 \times 10^{-4} \text{ s}^{-1}$, equivalent to a macroscopic half-life of 30 min. The operators are sufficiently positioned so that DNA looping may not occur.

With only a single operator, 29.8% of the trials have a certainty of 0.8 or greater, including both the on and off

states. By inserting additional operators, the certainty of the bistable switch increases. With two and three operators, 42.6 and 43.7% of the trials, respectively, have a certainty of 0.8 or greater. The addition of the third operator grants only a small increase in the certainty of the switch, indicating that complete transcriptional control may not be achieved by simply adding more operators.

3.3. The effect of DNA looping

The ability for the repressor to bind to two operators at once allows the repressor to form a stable complex, increasing transcriptional control. The effect of DNA looping is examined by varying the dissociation constant of the DNA loop complex (Fig. 6). The increase in effective DNA looping may be accomplished by specific positioning of the operators with respect to one another. Three computational experiments were performed where DNA looping was absent, or DNA looping was present with dissociation constants of 1.0×10^3 or 1.0 s^{-1} . Where DNA looping is present, loop formation occurred at $5 \times 10^3 \text{ M}^{-1} \text{ s}^{-1}$, resulting in equilibrium constants of 5.0 or $5 \times 10^3 \text{ M}^{-1}$, respectively. Only the first operator is capable of binding repressor in solution, whereas the second operator engages only in DNA looping, if present. The first operator has a dissociation constant of $3.85 \times 10^{-4} \text{ s}^{-1}$ when bound to a repressor.

DNA looping increases the apparent half-life of the repressor-DNA complex by forcing the repressor to unbind twice before being able to diffuse away from the DNA. With each decrease in the probability of loop separation, the transcriptional control increases, resulting in greater certainty. Designs utilizing DNA looping with equilibrium constants of 5.0 and $5.0 \times 10^3 \text{ M}^{-1}$, respectively, result in 34.8% and

37.5% of the trials having a certainty of 0.8 or greater. Additional increases in loop stability continue to augment the certainty of the switch's state. With greater DNA loop stability, once a repressor binds to the first operator, the probability of successful gene expression lessens dramatically. Because the loop structure requires only the presence of a single repressor, transcriptional control behaves more like a "hard switch", completely turning off gene expression as soon as a single repressor has bound.

3.4. The usage of self-repression to decrease steady state concentrations

In order to decrease the concentration of both repressor and co-expressed protein at steady state, one may apply self-repression, or autoregulatory negative feedback, by the insertion of a fourth operator into both genes. While the first through third operators bind to repressor produced by the opposing gene, the fourth operator binds to repressor produced by its own gene. Increasing the affinity of the fourth operator for the gene's product repressor will decrease the probability of successful transcription at high concentrations of repressor. Simulations are run where the fourth operator is non-existent or, when bound to a repressor, has a disassociation constant of 3.85×10^{-3} or $7.7 \times 10^{-4} \text{ s}^{-1}$, equivalent to macroscopic half-lives of 3 or 15 min, respectively.

Initially, each design produces repressor at the same rate. However, as the concentration of repressor rises, each design reduces its production of repressor corresponding to the affinity of the repressor for the fourth operator (Fig. 7). At the end of 5 h of simulated time, the application of self-repression decreases the average concentration of repressor (solid lines)

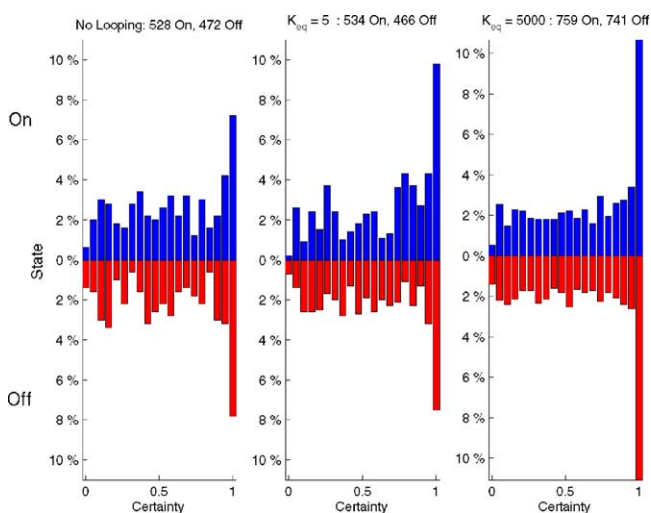


Fig. 6. A state vs. certainty histogram of three designs, each including DNA looping to a different extent. The first design excludes DNA looping, while the second and third utilize DNA looping with equilibrium constants of 5.0 and $5.0 \times 10^3 \text{ M}^{-1}$, respectively. All three designs contain only one operator capable of binding repressor in solution with a repressor-operator dissociation constant of $3.85 \times 10^{-4} \text{ s}^{-1}$.

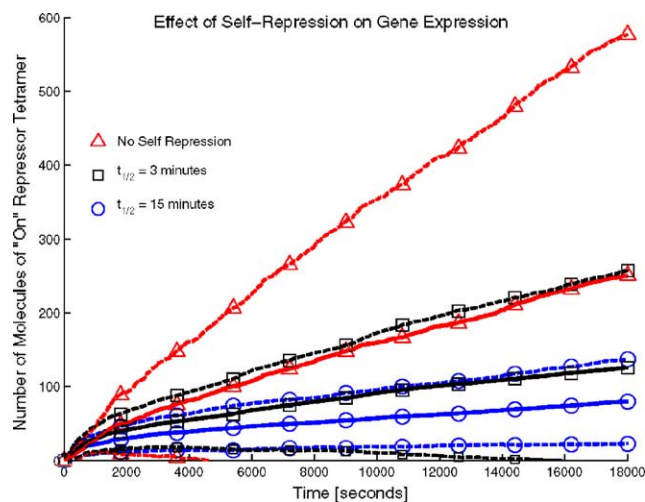


Fig. 7. The mean (solid lines) and ± 1 S.D. (dashed lines) of the number of molecules of R_2 , the "On" repressor, are plotted over time with varying degrees of self-repression. The first design excludes the fourth operator, applying no self-repression (triangles), while the second and third design includes the fourth operator with repressor-operator dissociation constants of 3.85×10^{-3} and $7.7 \times 10^{-4} \text{ s}^{-1}$, equivalent to macroscopic half-lives of 3 min (squares) and 15 min (circles), respectively.

by 50.0% and 68.2%, respectively, for repressor–operator complexes whose half-lives are 3 and 15 min. The application of self-repression also reduces the fluctuations in production of repressor and co-expressed protein (dotted lines). Because high concentrations of repressor lessen the probability of successful gene expression, the rate of production has an upper limit, decreasing the amount of fluctuations that might spur higher than average concentrations of protein. The fano factor, or the standard deviation divided by the mean (σ/μ), describes the extent of fluctuations in a series of measured values. The fano factor is 1.436 with no self-repression, decreasing to 1.08 and 0.699 with repressor–operator complex half-lives of 3 and 15 min, respectively. A fano factor of zero indicates the absence of fluctuations.

4. Discussion and conclusion

4.1. Precise regulation

The distribution of outcomes represents the possible states, and their certainties, that a colony of cells, all containing an implementation of a bistable switch, may produce. In biological systems, the faulty performance of a small number of cells may have significant consequences to the health of the organism. In the previous results, there are considerable instances where cells produce faulty switches, ones with low certainty. One reason for this lack of precision is the absence of translational and post-translational regulation. It may be that the need for precise regulation is the primary evolutionary force for developing new regulatory mechanisms, those that are common in eukaryotes and higher organisms. The exclusive application of transcriptional regulation, modifying the efficiency of successful mRNA production, cannot guarantee the necessary redundancy and precision that is required by higher organisms to regulate their most critical processes. In order to evolve more complicated processes, additional regulatory mechanisms must evolve simultaneously in order to ensure the viability of the organism.

4.2. Design principles

A high quality bistable switch design requires tight transcriptional control and strong self-repression to obtain rapid switching between discrete states. A combination of multiple operators and stable loop formation provides the transcriptional control necessary to distinctly separate the on and off states and give a high certainty of differentiation. The application of self-repression decreases the steady state concentration of repressor and co-expressed protein of the dominant state. When induced, the concentration of the dominant repressor will go to zero faster, increasing the speed in which the switch may change states.

4.2.1. The differences between transcriptional control elements

The usage of multiple operators and the formation of stable loops provide two different forms of transcriptional control. Designs with multiple operators provide increased sensitivity to the concentration of repressor in solution. The probability of a repressor binding to any of the operators is a direct function of the concentration of the repressor. As more operators are added, there is an additive probability that at least one operator binds to a repressor in solution. Even when bound, the operators still maintain sensitivity to the concentration of repressor in solution because, over time, the operators will unbind from their repressor and the probability of being rebound will depend on the concentration in solution. In contrast, stable loop formation only requires the presence of a single bound repressor and is not sensitive to the concentration of repressor in solution. Once formed, the stable loop complex has a small probability of full dissociation, because the repressor must unbind twice in order to be free. While formed, the loop's continued presence is not a strong function of the concentration of repressor in solution. In order to activate gene expression, another participatory molecule, such as an inducer, must be present to disrupt the repressor's ability to form the loop complex, consequently ceasing repression.

The two disparate types of transcriptional control elements have different uses. For the bistable switch, one desires a "hard" immovable switch, sensitive to the environment for only the first few moments after induction and thereafter static. The formation of stable DNA loop structures provides this effective characteristic. For other designs, however, a static transcriptional mechanism may not be desired. The usage of multiple operators, then, becomes the more effective transcriptional control mechanism for a dynamic environment.

4.2.2. Self-repression reduces the time necessary to switch from one state to another, but also increases the likelihood that neither gene will be expressed

The application of self-repression by the inclusion of an operator which binds to the gene's own product decreases the steady state concentration of the gene's expressed proteins. When an inducer is added to the system, the repressor of the dominantly expressed gene binds less specifically to the opposing genes' operators. However, because the activity of the repressor depends on both specificity and concentration, the dominantly expressed gene will maintain dominance if its repressor is at a sufficiently high concentration. Self-repression reduces the probability that the dominant genes' repressor maintains a high concentration and increases the probability that the opposing gene will become expressed when inducer is added. In order for a measurable state change to occur, the opposing gene must become expressed, producing its repressor and co-expressed protein, the dominant gene must be repressed, ceasing production of its proteins, and the dominant genes' proteins in solution must degrade rapidly.

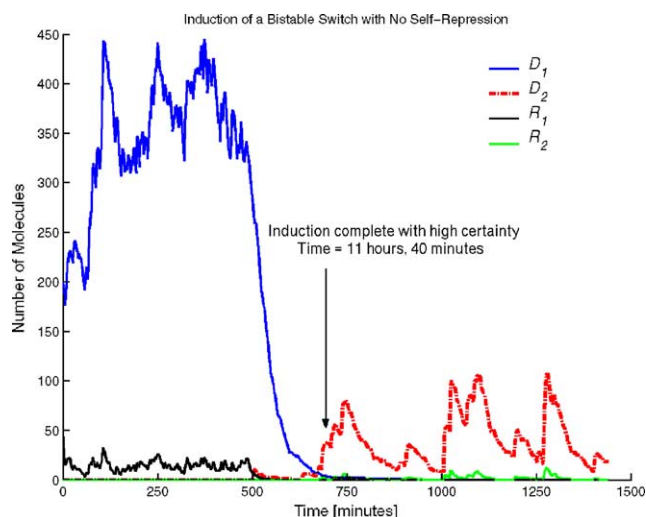


Fig. 8. One possible outcome of the induction of a quality bistable switch design, excluding any self-repression. The number of molecules of co-expressed protein, D, and repressor, R, from both genes one and two are plotted vs. time. The design consists of two operators, both capable of binding repressor in solution and forming a DNA loop complex with one another. Repressor–operator complexes have a dissociation constant of $3.85 \times 10^{-4} \text{ s}^{-1}$ and DNA loop formation has an equilibrium of $5.0 \times 10^3 \text{ M}^{-1}$. Initial conditions were taken from near-steady state concentrations of an “Off” state using the same kinetic parameters.

Self-repression reduces the steady state concentration of the dominant genes’ proteins, decreasing the time necessary for complete degradation, and consequently increasing the rate of state switching (see Figs. 8 and 9). In both figures, induction is initiated at time zero by adding the “On” inducer, which alters the ability of the “Off” repressor to bind to its DNA operators. Induction is considered complete when the number of “On” co-expressed protein, D_2 , is much greater than the “Off” co-expression protein, D_1 . Without self-repression, the large concentration of “Off” repressor, R_1 , increases the length of time necessary for induction to complete.

However, excessive self-repression results in the steady state concentration of the dominant genes’ proteins nearing zero. With the auto regulating operator too “tight”, small amounts of repressor will prevent any occurrences of successful transcription, resulting in rare and discontinuous production of mRNA. If the design includes tight self-repression in both genes, then neither gene will become sufficiently expressed to produce a discrete state of the switch, resulting in a near zero certainty (results not shown).

4.3. Conclusions

The bistable switch gene circuit will prove to be a highly useful motif in future biological applications, enabling rapid, programmed responses to specific intracellular stimuli and producing decision-making branch points in the “programming” of biological organisms. The design of the bistable switch is sensitive to the number, affinity, and placement of the operators, giving genetic circuit engineers a great deal of

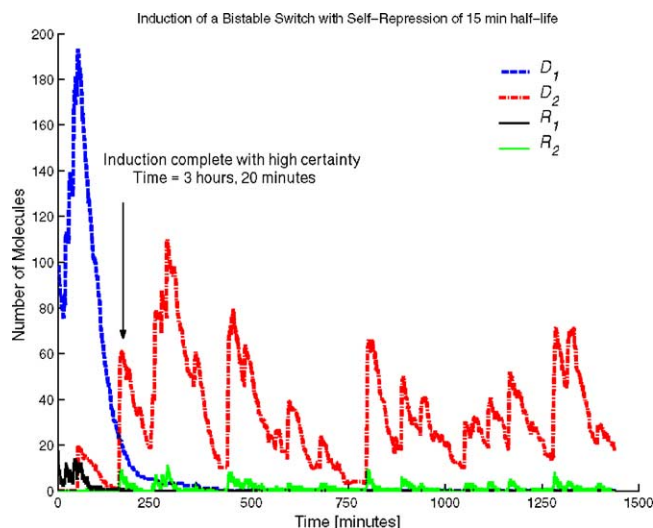


Fig. 9. One possible outcome of the induction of a quality bistable switch design, including self-repression with a repressor–operator dissociation constant of $7.7 \times 10^{-4} \text{ s}^{-1}$, equivalent to a half-life of 15 min. The number of molecules of co-expressed protein, D, and repressor, R, from both genes one and two are plotted vs. time. The design consists of two operators, both capable of binding repressor in solution and forming a DNA loop complex with one another. Repressor–operator complexes have a dissociation constant of $3.85 \times 10^{-4} \text{ s}^{-1}$ and DNA loop formation has an equilibrium of $5.0 \times 10^3 \text{ M}^{-1}$. Initial conditions were taken from near-steady state concentrations of an “Off” state using the same kinetic parameters.

flexibility in producing a high quality design. Using multiple operators, DNA looping, and self-repression, one can create a bistable switch that rapidly switches between discrete states.

As more biological systems become well characterized, the number of usable “building blocks” will increase, greatly profiting the design of future gene circuits. As more gene circuits are quantitatively analyzed, they may be combined to create more complex functions that add to the genetic circuit engineer’s toolbox. The synthesis of genetics, molecular biology, and chemical engineering will open up a new area of design, one where we specifically manipulate the genomes of biological organisms to perform a wide array of useful applications.

References

- Ansari, A. Z., Mapp, A. K., Nguyen, D. H., Dervan, P. B., & Ptashne, M. (2001). Towards a minimal motif for artificial transcriptional activators. *Chemistry and Biology*, 8, 583–592.
- Arkin, A. P., Ross, J., & McAdams, H. H. (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells. *Genetics*, 149, 1633–1648.
- Beckwith, J. R., & Zipser, D. (Eds.). (1970). *The Lactose Operon*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.
- Borukhov, S., & Nudler, E. (2003). RNA Polymerase holoenzyme: structure, function, and biological implications. *Current Opinion in Microbiology*, 6, 93–100.
- Burgess, R. R., & Anthony, L. (2001). How sigma docks to RNA Polymerase and what sigma does. *Current Opinion in Microbiology*, 4, 126–131.

- Coulombe, B., & Burton, Z. F. (1999). DNA bending and wrapping around RNA Polymerase. *Microbiology and Molecular Biology Reviews*, 63(2), 457–478.
- Dai, X., & Rothman-Denes, L. B. (1999). DNA structure and transcription. *Current Opinion in Microbiology*, 2, 126–130.
- Elowitz, M. B., & Leibler, S. (20 January 2000). A synthetic oscillatory network of transcriptional regulators. *Nature*, 403, 335–338.
- Franch, T., & Gerdes, K. (2000). U-turns and regulatory RNAs. *Current Opinion in Microbiology*, 3, 159–164.
- Friedman, A. M., Fischmann, T. O., & Steitz, T. A. (23 June 1995). Crystal of *Lac* repressor core tetramer and its implications for DNA looping. *Science*, 268, 1721–1727.
- Gardner, T. S., Cantor, C. R., & Collins, J. J. (20 January 2000). Construction of a genetic toggle switch in *Escherichia coli*. *Nature*, 403, 339–342.
- Gibson, M. A., & Bruck, J. (2000). Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Physical Chemistry A*, 104, 1876–1889.
- Gillespie, D. T. (1976). A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *Journal of Computational Physics*, 22, 403–434.
- Gillespie, D. T. (1992). *Introduction to Markov processes for physical scientists*. Boston: Academic Press.
- Hasty, J., McMillen, D., & Collins, J. J. (14 November 2002). Engineering gene circuits. *Nature*, 420, 224–230.
- Jenal, U., & Hengge-Aronis, R. (2003). Regulation by proteolysis in bacterial cells. *Current Opinion in Microbiology*, 6, 163–171.
- Kercher, M. A., Lu, P., & Lewis, M. (1997). *Lac* repressor–operator complex. *Current Opinion in Structural Biology*, 7, 76–85.
- Levandoski, M. M., Tsodikov, O. V., Frank, D. E., Melcher, S. E., Saecker, R. M., & Record, M. T., Jr. (1996). Cooperative and anticooperative effects in binding of the first and second plasmid O^{sym} operators to a *LacI* tetramer: Evidence for contributions of non-operator DNA binding by wrapping and looping. *Journal of Molecular Biology*, 260, 697–717.
- Masse, E., Majdalani, N., & Gottesman, S. (2003). Regulatory roles for small RNAs in bacteria. *Current Opinion in Microbiology*, 6, 120–124.
- McKnight, S. L., & Yamamoto, K. R. (Eds.). (1992). *Transcriptional regulation*. Plainview: Cold Spring Harbor Laboratory Press.
- McLeod, S. M., & Johnson, R. C. (2001). Control of transcription by nucleoid proteins. *Current Opinion in Microbiology*, 4, 152–159.
- Müller, J., Oehler, S., & Müller-Hill, B. (1996). Repression of *lac* promoter as a function of distance, phase, and quality of an auxiliary *lac* operator. *Journal of Molecular Biology*, 257, 21–29.
- Orphanides, G., & Reinberg, D. (2002). A unified theory of gene expression. *Cell*, 108, 439–451.
- Ptashne, M. (1996). *A genetic switch: gene control and phage lambda*. Cambridge, MA: Cell Press and Blackwell Scientific Publications.
- Ramakrishnan, V. (2002). Ribosome structure and mechanism of translation. *Cell*, 108, 557–572.
- Rojo, F. (2001). Mechanisms of transcriptional repression. *Current Opinion in Microbiology*, 4, 145–151.
- Romby, P., & Springer, M. (2003). Bacterial translational control at atomic resolution. *Trends in Genetics*, 19(3), 155–161.
- Sorensen, M. A., & Pedersen, S. (1991). Absolute in vivo translation rates of individual codons in *Escherichia coli*. *Journal of Molecular Biology*, 222, 265–280.
- Vogel, U., & Jensen, K. F. (1994). The RNA chain elongation rate in *Escherichia coli* depends on the growth rate. *Journal of Bacteriology*, 176, 2807–2813.
- Wolf, D. M., & Arkin, A. P. (2002). Fifteen minutes of *fim*: Control of type 1 pili expression in *E. coli*. *OMICS: A Journal of Integrative Biology*, 6(1), 91–114.
- Wolf, D. M., & Arkin, A. P. (2003). Motifs, modules, and games in bacteria. *Current Opinion in Microbiology*, 6, 125–134.
- Xu, H., & Hoover, T. R. (2001). Transcriptional regulation at a distance in bacteria. *Current Opinion in Microbiology*, 4, 138–144.