Model-Driven Designs of an Oscillating Gene Network

Lisa M. Tuttle, Howard Salis, Jonathan Tomshine, and Yiannis N. Kaznessis
Department of Chemical Engineering and Materials Science, and Digital Technology Center, University of Minnesota, Minneapolis, Minnesota 55455

ABSTRACT The current rapid expansion of biological knowledge offers a great opportunity to rationally engineer biological systems that respond to signals such as light and chemical inducers by producing specific proteins. Turning on and off the production of proteins on demand holds great promise for creating significant biotechnological and biomedical applications. With successful stories already registered, the challenge still lies with rationally engineering gene regulatory networks which, like electronic circuits, sense inputs and generate desired outputs. From the literature, we have found kinetic and thermodynamic information describing the molecular components and interactions of the transcriptionally repressing *lac*, *tet*, and *ara* operons. Connecting these components in a model gene network, we determine how to change the kinetic parameters to make this normally nonperiodic system one which has well-defined oscillations. Simulating the designed *lac-tet-ara* gene network using a hybrid stochastic-discrete and stochastic-continuous algorithm, we seek to elucidate the relationship between the strength and type of specific connections in the gene network and the oscillatory nature of the protein product. Modeling the molecular components of the gene network allows the simulation to capture the dynamics of the real biological system. Analyzing the effect of modifications at this level provides the ability to predict how changes to experimental systems will alter the network behavior, while saving the time and expense of trial and error experimental modifications.

INTRODUCTION

Gene regulatory networks are important for sensing a wealth of molecular signals in a biological system and for generating a specific output that has either naturally evolved to increase the fitness of the organism or has been designed by engineers to serve a specific function (1,2). However, it is still not well understood how interactions at the molecular level influence the system's dynamic behavior. To understand how these interactions affect the protein production of the system, it is essential that models are developed that account for all the known components at the molecular level. Specifically, rational engineering of such inducible gene networks requires intervention at two levels:

1. The level of network connections, where DNA binding proteins regulate the expression of specific genes by either activation or repression. By combining simple regulatory interactions, such as negative and positive feedback and feed forward loops, one may create more intricate networks that precisely control the production of protein molecules, such as bistable switches, oscillators, and filters. In the laboratory, these networks can be created using existing libraries of regulatory proteins and their corresponding operator sites.
2. The level of molecular components, which describes the kinetics and strengths of the protein-DNA, protein-RNA, and protein-protein interactions within the system. The dynamical behavior of the system is a complex function of the kinetic interactions of the components. By altering the characteristics of the components, such as DNA-binding proteins and their corresponding DNA sites, one can modify the system's dynamical behavior without modifying the network level connections. In the laboratory, the DNA sequences that yield the desired characteristics of each component can be engineered to achieve the desired protein-protein, protein-RNA, or protein-DNA binding constants and enzymatic activities.

The large number of components and interactions involved in dynamic gene regulation requires computational modeling, since the cost of experimentally changing these components and the kinetics of their interaction is large. Computer simulations enable exhaustive searches of different network connectivities and molecular thermodynamic/kinetic parameters, greatly advancing the development of simple rules, or design principles, that seek to simplify the complicated behavior of the network into a brief, usable framework. In this work, we present design rules for constructing a robust oscillating gene network, or repressillator. Our objective is to quantify the effects of the molecular level interactions on the period and variability of the oscillating protein concentrations. Oscillatory gene networks are prevalent in nature, especially ones that generate circadian rhythms (3,4). Given the naturally oscillatory levels of many chemicals in the body, it will be useful to have well-controlled oscillating protein levels for use in applications such as chronopharmaceutics (5).

Previous work has studied naturally oscillating systems using deterministic models and techniques from the study of nonlinear dynamics, including bifurcation analysis. The

Submitted April 7, 2005, and accepted for publication September 12, 2005.

Address reprint requests to Yiannis Kaznessis, Dept. of Chemical Engineering and Materials Science, University of Minnesota, 421 Washington Ave. SE, Minneapolis, MN 55455. Tel.: 612-624-4197; Fax: 612-626-4276; E-mail: yiannis@cems.umn.edu.

0006-3495/05/12/3873/11 $2.00

doi: 10.1529/biophysj.105.064204

Drosophila circadian rhythm (6), a simplified model of circadian rhythm combining positive and negative regulation (7,8), and the entrainment of a synthetic oscillator to a bacterial cell cycle (9) have been mathematically modeled and analyzed. However, deterministic methods make multiple approximations on the continuity and differentiability of the reaction events that occur within a biological system. These approximations have been shown to be invalid for many biological processes, especially gene expression (10). Here, we describe the system's dynamics using a fully stochastic representation and use stochastic simulations to compute an ensemble of trajectories.

Following the work of Elowitz and Leibler (1), we connect three genes in a cycle of negative feedback loops. The protein products repress the expression of the gene next in sequence, creating the possibility of sustained oscillations. We quantitatively model the repressilator by extracting components from the well-characterized *lac*, *tet*, and *ara* operons, using kinetic parameters from the literature (11–17), and inserting them in the network configuration of the repressilator. Besides the natural components, a wide variety of *lac* and *tet* DNA sites and repressor proteins have been created through extensive mutatagenesis, each with altered kinetic characteristics (18–21).

There are two main differences between our approach and previously developed models. We use a detailed, mechanistic model of the components and interactions that constitute bacterial transcription and translation, including transcriptional and translational elongation and the binding of proteins to individual DNA sites. We include all protein-protein interactions, including dimerization and tetramerization reactions. We do not simplify the model by reducing multiple biomolecular interactions to a few functional groups (in an extreme example we could use a model in which DNA produces RNA which produces protein, using phenomenological transcription and translation kinetic constants). We describe the processes of transcriptional and translational elongation as a Γ-distributed event with the rates of elongation of 30 nucleotides per second and 33 amino acids per second, respectively. We also simulate the dynamics of the network using a hybrid stochastic-discrete and stochastic-continuous algorithm. The models capture the behavior of single cells more accurately than deterministic kinetics, which require that the system be at the thermodynamic limit. Since species such as promoter or operator sites may only be present in single molecule quantities, their concentrations may not be modeled as continuous. Instead, our hybrid stochastic algorithm correctly simulates a coupled jump and continuous Markov process describing, respectively, the dynamics of the stochastic-discrete reactions and the stochastic-continuous reactions. A purely Langevin approach would incorrectly approximate the discrete-stochastic reactions, such as repressor proteins binding to individual DNA sites. It is rather straightforward to develop a deterministic-continuous model of biomolecular interactions that will capture existing experimentally measured concentration profiles. It is more challenging to use a model with sufficient detail, in terms of specific molecular species and interactions, to generate design rules for new gene regulatory networks which can then be directly tested in the lab. The disadvantage of our approach is that the mathematical analysis of the dynamics of the system is numerically performed and requires a large number of simulations. Further, in contrast to deterministic nonlinear dynamics, the theory behind the stability and bifurcation analysis of discrete or continuous stochastic systems is far less developed. To offset this disadvantage, we use techniques from the design of electronic circuits, such as the cyclic covariance function, to compute the periods of stochastic limit cycles.

In this work, we propose multiple configurations of network connectivities and their kinetic parameters that result in a robust repressilator. We begin the design work with two configurations, each creatable from currently existing molecular parts, and show that these configurations do not result in sustained oscillations. We then pinpoint multiple mutations which will exhibit more robust oscillations. Finally, we perform a sensitivity analysis of each design parameter of the system, showing the effects of the number of operators, the operator-repressor affinities, the mRNA and protein half-lives, and the numbers of ribosome and RNA polymerase on the period of oscillation. The information should be useful to any synthetic biologist hoping to construct an oscillating gene network.

MODELS AND METHODS

Hybrid stochastic simulation algorithm

For a system of N species S_i ($i = 1,\ldots,N$) reacting via M reaction pathways $\mathfrak{R}_j$ ($j = 1,\ldots,M$) in a volume V we define the following:

- The state vector of the system as the N-dimensional vector $\boldsymbol{X}(t) = \{X_1(t),\ldots, X_N(t)\}$, where X_i is the number of molecules of species S_i at time t.
- The $M \times N$ stoichiometric reaction matrix, $\boldsymbol{\nu}$, where ν_{ij} is the change in the number of S_i molecules produced by a reaction $\mathfrak{R}_j$.
- $a_j(\boldsymbol{x})dt$ = the probability that a reaction $\mathfrak{R}_j$ will occur somewhere in the system volume in the time interval $[t, t + dt]$. We can write $a_j = h_j k_j$, where h_j is the number of possible combinations of the reacting molecules in $\mathfrak{R}_j$ and k_j is the reaction rate constant.

The original stochastic simulation algorithm (SSA) of Gillespie (22) exactly simulates trajectories of a jump Markov process described by the Master equation. Improved variants have incrementally decreased the computational cost (23,24) while retaining the jump Markov description. However, because the computational cost scales with the number of reaction occurrences, systems with one or more ''fast'' reactions become costly to simulate as a jump Markov process. In the case of the repressilator, many biomolecular interactions, especially the protein dimerization ones, are considered ''fast'' and require extensive computational time to execute their individual events. By assuming these fast reactions occur continuously, we can convert their mathematical representation to a continuous Markov process and describe their dynamics with a system of chemical Langevin equations (CLEs) (25). The result is a system of Itô stochastic differential equations (SDEs) with multiple multiplicative noises, or

$$dX_i = \sum_{j=1}^{M^{fast}} \nu_{ji} a_j^f(\boldsymbol{X}(t))dt + \sum_{j=1}^{M^{fast}} \nu_{ji} \sqrt{a_j^f(\boldsymbol{X}(t))} dW_j, \quad (1)$$

where a^f is the fast reaction propensities, W is an M^{fast} dimensional Wiener process, and ν is correspondingly altered to include only the fast reactions.

The challenge lies with integrating a discrete-stochastic model, like the SSA, with a continuous stochastic model, like the CLE. We have recently developed a hybrid stochastic algorithm that combines the two and significantly outperforms the SSA while retaining its accuracy (26). The algorithm partitions the system into subsets of fast/continuous and slow/discrete reactions, uses the CLE to describe the effects of the fast reactions, and solves for slow reaction times with a system of differential Jump equations, which are

$$dR_j(t) = a_j^s(\boldsymbol{X}(t))dt \quad R_j(t_o) = \log(URN_j), \quad (2)$$

where R_j is called the reaction residual and URN is a uniform random number between zero and one. The differential Jump equations are solved by randomly selecting their negative initial conditions, integrating them forward in time, and monitoring the zero crossings of the reaction residuals. The jth slow reaction occurs at time τ_j when $R(\tau_j) = 0$. The system of Jump equations are also SDEs because they are coupled to the CLE via the state vector, $\boldsymbol{X}(t)$. By using a stochastic numerical integrator, such as the Euler-Maruyama or Milstein methods, one can solve the coupled system of SDEs and determine the global error of both the CLEs and the slow reaction times. The method has been shown to be accurate and can be many orders of magnitude faster than the original SSA when one or more fast reactions exist. We use this hybrid stochastic simulation method to compute the stochastic dynamics of the repressilator gene network. For each set of kinetic parameters, at least 100 independent trajectories are computed.

Quality of oscillations

Since gene expression is a stochastic process, single cells may exhibit widely different behaviors, even if they somehow begin with the same initial conditions. Because of internal noise, oscillations in protein or mRNA molecules will have fluctuating periods, amplitudes, and phases. To make these oscillations useful for some purpose, the gene network must be designed so as to minimize the fluctuations in the period and amplitude. To quantitatively characterize the stochasticity of oscillations, we use a method taken from the design of electronic circuits. We assume the oscillating protein signals are a cyclostationary signal and use the Fourier transform of their autocorrelation functions to compute the average and standard deviation of the periods of oscillation. The method works well even when an oscillating protein signal is partially masked by background stochasticity.

The oscillatory concentrations of the molecules of species S_k are described as a cyclostationary signal. With $X_k(t)$ being a discrete-index random process, we can define the mean of this time series as $\mu_x(t) := E\{X_k(t)\}$ and the covariance $c_{xx}(t;\tau) := E\{[X_k(t) - \mu_x(t)][X_k(t+\tau) - \mu_x(t+\tau)]\}$. The signal $X_k(t)$ is then called cyclostationary if there exists an integer q such that $\mu_x(t) = \mu_x(t + lq)$ and $c_{xx}(t;\tau) = c_{xx}(t + lq;\tau)\ \forall t, l \in \mathbb{Z}$. To best determine the period of oscillation, the cyclic correlation function, C_{xx}, can be computed as follows (27)

$$C_{xx}(\alpha;\tau) = \frac{1}{T}\sum_t \{[X_k(t) - \mu_x] \times [X_k(t+\tau) - \mu_x]\}\, e^{-j\alpha t}, \quad (3)$$

where T is the number of data points, τ is taken to be 0, and j is the square root of -1. The cycle parameter, α, is $2\pi n/P$, where P is the period of oscillation and n is an integer number. A plot of C_{xx} versus α gives peaks corresponding to the most dominant periods of oscillation in the signal $X_k(t)$. There will always be a dominant peak at $\alpha = 0$ because n may be 0, representing the infinite period. For each simulation trial and each oscillating protein species, the dominant nonzero peak, corresponding to $n = 1$, is chosen and is used to determine the period. Additional peaks will appear in harmonic multiples, corresponding to larger values of n. The reported period is an average over all trials. The standard deviation of the period is based on the standard deviation of the α-values, such that $\sigma_P = 2\pi\sigma_\alpha/\alpha^2$, where the α used is the average value over the trials. The cyclic correlation curves shown in this work are obtained by averaging over all the trials and are normalized so that the highest amplitude is one. A system exhibiting sustained oscillations with few fluctuations in its period will produce a well-defined α-peak in the averaged cyclic correlation functions.

The *lac-tet-ara* gene network

The *lac-tet-ara* system is an experimentally realizable gene network with many possible combinations of individual molecular components. The network connections are constructed so that sustained oscillations are possible, but not guaranteed. The production of LacI monomers is repressed by $AraC_2$ proteins bound to promoter-overlapping I_1/I_2 sites, the production of TetR monomers is repressed by $LacI_4$ tetramers bound to one or more promoter-overlapping *lac* operators, and the production of AraC monomers is repressed by $TetR_2$ dimers bound to one or more promoter-overlapping *tet* operators.

The *lac*, *tet*, and *ara* operators may be moved, replicated with one or more adjacent copies, or replaced with mutant variants. The 5′ untranslated region (5′ UTR) of the repressor mRNAs may be altered to increase or decrease their degradation rates. The repressor proteins may also be fused with ssrA peptides to increase their degradation rates. One configuration, consisting of one wild-type operator regulating each gene and using wild-type repressor proteins and mRNAs, is shown in Fig. 1. Its corresponding mechanistic system of reactions is detailed in Table 1. Although there are

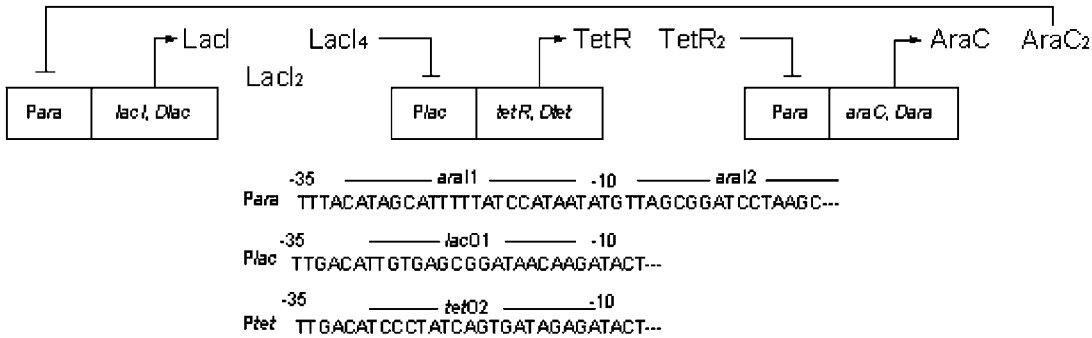

FIGURE 1 Network connectivity for a *lac-tet-ara* oscillating gene network. Below, the sequences of the promoter regions, using a single, promoter-overlapping operator per gene.

TABLE 1 Base reactions and kinetic rates for the *lac-tet-ara* system

Rxn	Reaction	k	Ref.
1	2 LacI → LacI2	1.0e9	7
2	LacI2 → 2 LacI	10	7
3	2 LacI2 → LacI4	1.0e9	7
4	4: LacI4 → 2 LacI2	10	7
5	LacI4 + lacO1 → LacI4:lacO1	5.e9	8
6	LacI4:lacO1 → LacI4 + lacO1	3.85e-4	8
7	2 tetR → tetR2	1.0e9	§
8	tetR2 → 2 tetR	10	§
9	tetR2 + tetO2 → tetR2:tetO2	2.98e6	11
10	tetR2:tetO2 → tetR2 + tetO2	2.13e-2	11
11	2 araC → araC2	1.0e9	§
12	araC2 → 2 araC	10	§
13	araC2 + araI1/I2 → araC2:araI1/I2	1.0e7	13¶
14	araC2:araI1/I2 → araC2 + araI1/I2	4.0e-3	13¶
15	RNAp + lacP:lacO1 → RNAp:lacP:lacO1	2.0e6	8
16	RNAp + tetP: tetO2 → RNAp:tetP:tetO2	8.6e5	12
17	RNAp + araP:araI1/I2 → RNAp:araP:araI1/I2	2.0e8	14¶
18	RNAp:lacP:lacO1 → RNAp:lacP*	0.01	8
19	RNAp:tetP:tetO2 → RNAp:tetP*	0.13	12
20	RNAp:araP:araI1/I2 → RNAp:araP*	0.167	14
21	RNAp:lacP* → lacP:lacO1 + RNAp:DNAlac	30 nt/s	9
22	RNAp:tetP* → tetP:tetO2 + RNAp:DNAtet	30 nt/s	9
23	RNAp:araP* → araP:araI1/I2 + RNAp:DNAara	30 nt/s	9
24	RNAp:tetP:tetO2 → RNAp + tetP:tetO2	0.10	12
25	RNAp:araP:araI1/I2 → RNAp + araP:araI1/I2	0.06	14¶
26	RNAp:lacP:lacO1 → RNAp + lacP:lacO1	0.01	8
27	RNAp:DNAlac → RNAp + tet_mRNA	30 nt/s, 660 nt	9
28	RNAp:DNAtet → RNAp + ara_mRNA	30 nt/s, 660 nt	9
29	RNAp:DNAara → RNAp + lac_mRNA	30 nt/s, 660 nt	9
30	lac_mRNA + rib → rib:lac_mRNA	1.0e5	†
31	tet_mRNA + rib → rib:tet_mRNA	1.0e5	†
32	ara_mRNA + rib → rib:ara_mRNA	1.0e5	†
33	rib:lac_mRNA → rib:lac_mRNA_1 + lac_mRNA	33 aa/s	10
34	rib:tet_mRNA → rib:tet_mRNA_1 + tet_mRNA	33 aa/s	10
35	rib:ara_mRNA → rib:ara_mRNA_1 + ara_mRNA	33 aa/s	10
36	rib:lac_mRNA_1 → rib + LacI + Dlac	33 aa/s, 220 aa	10
37	rib:tet_mRNA_1 → rib + tetR + Dtet	33 aa/s, 220 aa	10
38	rib:ara_mRNA_1 → rib + araC + Dara	33 aa/s, 220 aa	10
39	LacI →	2.31e-3	1
40	tetR →	2.31e-3	1
41	araC →	1.93e-4	‡
42	Dlac →	3.85e-4	‡
43	Dtet →	3.85e-4	‡
44	Dara →	3.85e-4	‡
45	LacI2 →	2.31e-3	1
46	LacI4 →	2.31e-3	1
47	tetR2 →	2.31e-3	1
48	araC2 →	1.93e-4	‡
49	lac_mRNA →	2.0e-3	†
50	tet_mRNA →	2.0e-3	†
51	ara_mRNA →	2.0e-3	†

Units on k: first order reaction, s^{-1}; second order, $(M\ s)^{-1}$. Reactions with two kinetic constants are Γ-distributed events, where the first number is the rate of each step and the second is the number of steps.

*Activated species, so that the transcription can proceed.

†Values were adjusted to give ~20 proteins per mRNA.

‡Based on typical protein degradation half-lives.

§Values were estimated for *tet* and *ara* parameters based on literature values for the *lac* system.

¶The forward and backward reaction rates were estimated from a given K_d value.

many possible configurations, we will first focus on ones that are synthesizable from currently available molecular components, including ones using the inducible promoters created by Bujard (28,29) and another using a simpler, single operator design.

When replacing wild-type DNA sites, proteins, or mRNA molecules with mutant variants, we modify the system of reactions with the altered kinetics. However, if we replicate one or more copies of an operator and place them adjacent to an existing one, we instead add more reactions to the system, including the new operators' interactions with repressors and RNA polymerases. A brief list of experimentally well-characterized mutant DNA sites and repressor proteins from the *lac* and *tet* operons is shown in Table 2. By using a detailed mechanistic model, we can accurately simulate the effects of adding new genetic elements without making possibly invalid approximations.

To connect the simulation results to experimentally observable phenotypes, the coding sequence for a fluorescent protein, such as green fluorescent protein (GFP), yellow fluorescent protein (YFP), or cyan fluorescent protein (CFP), is bicistronically added after each repressor coding sequence. The production of each fluorescent protein will be under the same control as the corresponding repressor and their concentration may be quantitatively measured, at the single cell level, with optical microscopy. In the model, the half-lives of the fluorescent proteins, which are also fused with the ssrA peptides, are kept constant at 30 min. For future applications, the fluorescent proteins may be replaced with other functional proteins, such as enzymes. We will refer to the fluorescent proteins as Dlac, Dtet, and Dara, referring to the repressor protein with which they are coexpressed.

Assumptions

There are a number of assumptions in the presented mathematical model. The reaction volume is considered to be a homogenous, well-stirred medium. The velocity distributions of all species must be a Maxwell-Boltzmann distribution after each reaction, requiring that many of the species' collisions are nonreacting and only aid in reaching the well-stirred approximation. Furthermore, it is assumed that other genes expressed in the reaction volume do not interfere with the studied gene networks, or that any of these effects are accounted for in the experimentally measured kinetic parameters. It is also assumed that there is no significant DNA binding of the monomer forms of LacI, TetR, and AraC and of the dimer form of LacI. Additionally, the small molecules lactose, tetracycline, and arabinose, often associated with these systems, are not included in the model. The repressor proteins are always capable of binding their cognate DNA sites and, once bound, sterically prevent the RNA polymerase from binding the promoter region. Although araC is an activator protein in the wild-type operon, one may convert it to a repressor protein by positioning its DNA binding sites so that the bound $araC_2$ overlaps with the promoter region.

For all trials, the initial numbers of free and available molecules for RNA polymerases and ribosome are 270 and 900, respectively. All present promoter and operator sites are initialized at one molecule. All other present chemical species are not initially present. Cell division is a discrete event which occurs every 30 ± 4 min, generated according to a Gaussian distribution. The volume of the cell is initially 10^{-15} liters and is linearly increased. The system volume is then halved as the cell is assumed to split into two equal daughter cells. Except for those species involving DNA sites, RNA polymerase, and ribosome, the numbers of molecules in the system are halved. The reported concentrations are on a per cell basis.

TABLE 2 A brief list of previously tested experimentally variant DNA sites and repressor proteins from the *lac* and *tet* operons

lac operator:repressor variants (18)	*tet* operator:repressor variants (19)
O_{sym}:wt K_{eq} = 1e11 [M^{-1}]	wt:wt rβ* = 0%
O_{4a}:wt K_{eq} = 1.1e8 [M^{-1}]	O_{2T}:KA33 rβ = 28%
O_{5a}:wt K_{eq} = 5.8e8 [M^{-1}]	O_{wt}:TA27 rβ = 45%
O_{5c}:wt K_{eq} = 8.1e7 [M^{-1}]	O_{3C}:wt rβ = 68%
O_{4a5c}:wt K_{eq} = 5.2e6 [M^{-1}]	O_{5G}:wt rβ = 87%

*The expression levels of *tet* operator:repressor variants are measured in relative β-galactosidase activity (rβ), where 100% expression occurs when repressor is absent.

RESULTS AND DISCUSSION

Designs using wild-type kinetics do not oscillate

We begin the design of an oscillating gene network by using only wild-type molecular components. One configuration, shown in Fig. 1, consists of only a single operator regulating each gene and uses the wild-type kinetics for the *tet*O_2, *lac*O_1, and *ara*I_1/I_2 DNA sites as well as the repressor mRNAs and proteins. In Fig. 2, we show the dynamical behavior and the cyclic covariance functions of the fluorescent proteins Dlac, Dtet, and Dara. Clearly, there are no sustained oscillations. The Dlac and Dara indicator proteins are constitutively expressed, whereas the production of Dtet protein remains fully repressed. The repressing action of the lacI tetramers is too strong, whereas the repressing action of the araC dimers is too weak. The cyclic covariance functions for all three species exhibit the expected dominant peak at the center, representing the infinite period. For the Dlac and Dara species, the function quickly decreases in amplitude with an initial peak at 0.225 h^{-1} and harmonic peaks at integer intervals, corresponding to the simulation's end time of 27.7 h. The Dtet species has no dominant period but the infinite one, resulting in the bow shape arc. If the system produces sustained oscillations, the cyclic covariance function would exhibit a

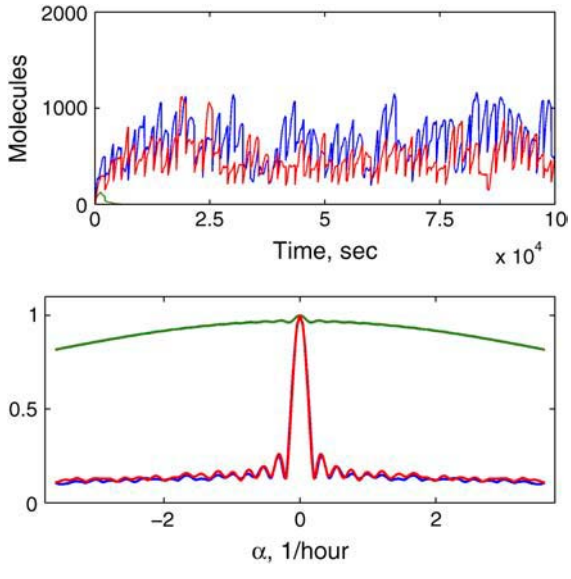

FIGURE 2 (*Top*) Dynamic behavior of the Dlac, Dtet, and Dara proteins in the simple example over 27.7 h. (*Bottom*) The cyclic covariance function of the same system.

sharp peak with an amplitude less than the central one but larger than the peaks corresponding to the simulation's end time.

In our next example configuration using only wild-type molecular components, we replace the regulatory regions from the simple example with a slightly modified version of the inducible promoter regions created by Lutz, Lozinski, Ellinger, and Bujard (28,29). The production of TetR monomers is regulated by $P_{LlacO-1}$, the production of AraC monomers is regulated by $P_{LtetO-1}$, and the production of LacI monomers is regulated by a modified P_{lar}. The promoter regions contain two *tet*O_2 operators, two *lac*O_1 operators, and a single *ara*I_1/I_2 site. In the P_{lar} promoter region, the *ara*I_1/I_2 sites are moved downstream so that the I_2 site is between the -35 and -10 consensus sequences and the I_1 site is adjacently upstream to the -35 region, causing the $AraC_2$ protein to act as a repressor, and the downstream *lac* operator is removed. The resulting dynamical behavior and the cyclic covariance functions of the indicator proteins Dlac, Dtet, and Dara are similar to the results of the first example (data not shown). The production of Dlac and Dara proteins is constitutive, whereas Dtet expression is fully repressed. The system exhibits no sustained oscillations, due to the same imbalance of repression between the three genes. We choose these two example designs for a specific purpose: computational modeling is able to quickly and cheaply discard hypothetical designs that will not produce the desired behavior. These systems would often be the first choices in an attempt to build an oscillating gene network, and we have shown that they do not function as intended.

An oscillating gene network with a 3-2-1 mutant operator configuration

We have created an asymmetric 3-2-1 operator design by modifying the Bujard promoter regions in two ways. The first is the creation of a promoter region containing three *lac* operators, combining the $P_{LlacO-1}$ and $P_{A1lacO-1}$ promoter regions, which regulate the production of TetR monomers. The *lac* and *tet* operators are also replaced with mutant variants. The three *lac* operators now have a decreased affinity to the $LacI_4$ tetramer with a K_{eq} of 5.2e11 [M^{-1}] and a dissociation kinetic constant of 1.93e-3 [s^{-1}], which is similar to the *lac*O_{sym} variant. The two *tet* operators now have an increased affinity to the $TetR_2$ repressor with a K_{eq} of 1.4e10 [M^{-1}] and a dissociation kinetic constant of 2.13e-2 [s^{-1}]. The half-life of the tetR protein is also decreased to 5 min. The *ara*I_1/I_2 site is kept as the wild type. The resulting dynamical behavior over a time interval of 5.8 days is shown in Fig. 3. There are only sustained oscillations in the Dara species and, to a lesser extent, in the Dtet species. To improve the quality of oscillations, the affinity of the *ara*I_1/I_2 site is increased to 5e10 [M^{-1}] with a dissociation kinetic constant of 4e-3 [s^{-1}] and the half-life of the TetR protein is reverted to its wild type. The resulting dynamical behavior

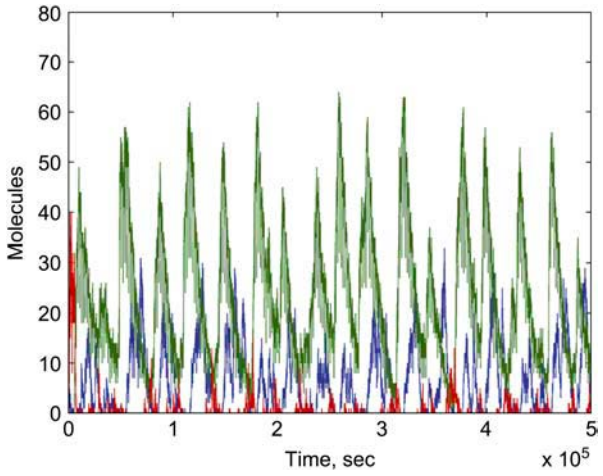

FIGURE 3 Dynamical behavior of the (*red*) Dlac, (*blue*) Dtet, and (*green*) Dara proteins using the 3-2-1 operator configuration with mutant *tet* and *lac* operators and a TetR protein half-life of 5 min.

and the corresponding cyclic covariance function are shown in Figs. 4 *A* and 5 *A*, respectively. Using the cyclic covariance functions, the average period of oscillations is computed as 16.2 h with a standard deviation of 4.1 h. The period of oscillations is much longer than the cell division time and is close to the period of a naturally occurring circadian rhythm. To further improve the quality of oscillations, the half-life of the TetR protein is again reduced to 10 min. The dynamical behavior of this last configuration and its corresponding cyclic covariance function is shown in Figs. 4 *B* and 5 *B*, respectively. The average period of oscillation decreases to 15.3 h with a standard deviation of 2.7 h. The amplitudes, or number of molecules at the peak of the oscillation, of all three fluorescent proteins are now roughly equal, and the variability in the period of the oscillation is decreased.

A systematic analysis of the oscillation envelope

Although such a heuristic approach will give useful results, the power of in silico modeling lies in the ability to conduct more exhaustive searches. To determine the width of the envelope of oscillation, a more thorough exploration of the parameter space is necessary. The initial concentrations of available RNA polymerases and ribosomes and the degradation rates of all proteins and mRNAs are varied. The effect of the number of operators and the repressor-operator affinity on the period of oscillation is also investigated. To isolate the effect of the particular variable under investigation, the asymmetric wild-type *lac*, *tet*, and *ara* operons were balanced by constructing a symmetric ''survey model'' in the same form as described in Table 1. This survey model contained between one and three operator sites regulating the production of each repressor. The model's rate constants are symmetric and consistent with the range of available molec-

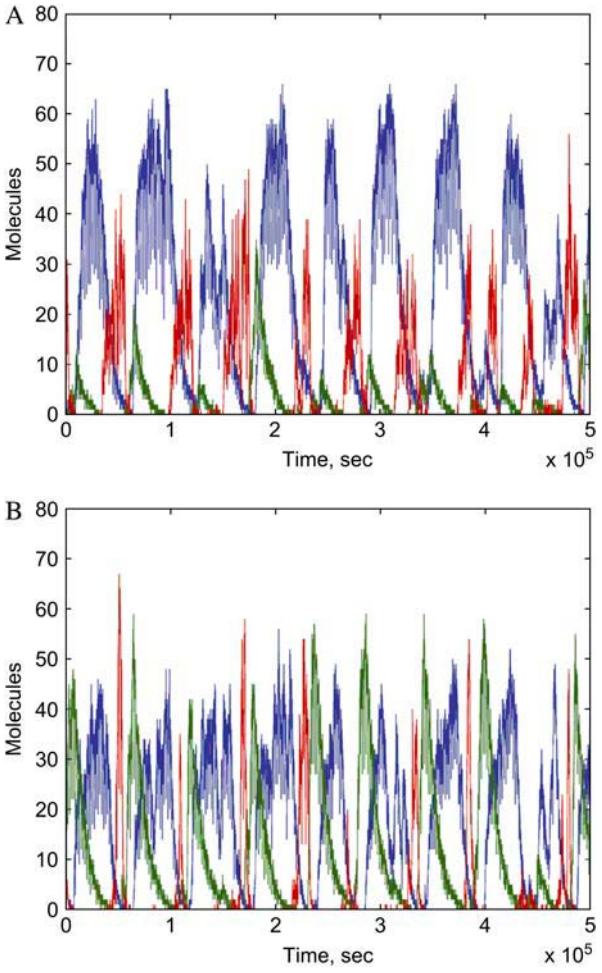

FIGURE 4 (*A*) Dynamical behavior of the (*red*) Dlac, (*blue*) Dtet, and (*green*) Dara proteins using a 3-2-1 operation design with mutant *tet*, *lac*, and *ara* operators and a TetR protein half-life of 30 min. (*B*) The same system as in *A*, but with a TetR protein half-life of 10 min.

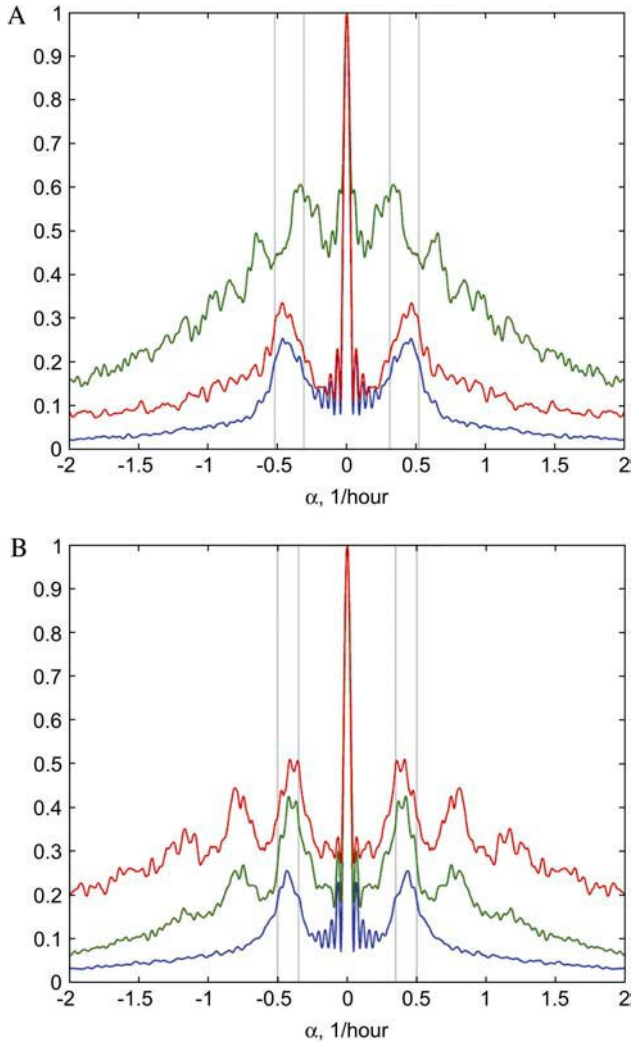

FIGURE 5 (*A*) Normalized average cyclic correlation functions of the (*red*) Dlac, (*blue*) Dtet, and (*green*) Dara proteins using the 3-2-1 operator configuration with mutant *tet*, *lac*, and *ara* operators and a TetR protein half-life of 30 min. The period of oscillation is 16.2 ± 4.1 h. (*B*) The same system as in *A*, but with a TetR protein half-life of 10 min. The period of oscillation is now 15.3 ± 2.7 h. The vertical gray lines represent the standard 68% confidence interval.

ular components, including both wild-types and mutant variants.

The affinity of the repressor for the operator sites is investigated by varying the equilibrium binding constants in the survey model through modification of the half-lives of the bound repressor-operator complexes. Since many DNA-protein associations have forward rate constants near the diffusion limit of $\sim 10^8$ $M^{-1}s^{-1}$ (30), this rate constant is fixed. The degradation rate of the complex was varied so as to give affinities between 10^{13} and 10^7 M^{-1}. This corresponds to half-lives between 19 h and 7 s. The results of the sensitivity analysis reveal the design rules for a three-gene repressilator.

The most basic of these rules is that total repression must be neither too strong, nor too weak. The symmetric survey model with three operator sites per regulated gene shows marked oscillations over a range of repressor-operator affinity of 10^9–10^{11} M^{-1}. The period of the oscillation also depends on this affinity—at 10^9 M^{-1}, the period is 3.39 h, at 10^{11} M^{-1}, the period increases to 11.55 h. When one operator site is removed from each gene, the resulting two-operator model shows marked oscillations over a somewhat wider range, giving periods from 20.04 h at an affinity of 10^{12} M^{-1} to 2.94 h at an affinity of 10^9 M^{-1}. With only a single operator per gene, the envelope of oscillation is very narrow near an affinity of 10^{11} M^{-1}, and the oscillations are irregular. Overall, higher repressor-operator affinity leads to a longer period of oscillation.

To investigate the effects of asymmetry, a similar series of models is constructed with the repressor-operator affinity fixed at 10^{10} M^{-1} for all of the operator sites of two genes, whereas the affinities of the sites on the third gene is varied.

When all genes have three operator sites, the asymmetric model oscillates over a somewhat wider range of affinities than the symmetric case, giving a period of 4.93 h at 10^9 M^{-1} (versus 3.39 h for the symmetric case) and 11.25 h at 10^{12} M^{-1} (versus no oscillations at this affinity in the symmetric case). With two active operator sites per gene, the model oscillates with periods from 3.85 to 7.78 h over this same range of affinities. With only a single operator site per gene, the system gives no regular oscillations. These results are summarized in Fig. 6.

These models show that a system may oscillate with repressor-operator affinities that differ by up to two orders of magnitude between the regulatory regions of each gene, provided that all genes contain the same number of operator sites. Excessive asymmetry in the amount of repression quenches the oscillations. The effect of increasing period with increasing affinity is also observed, although the trend is not as strong when the operator sites of only a single gene are varied. We observe that the oscillator's period is the sum of the pulse widths of all three fluorescent protein concentrations, where the pulse width is the amount of time required for each gene to express its repressor, become repressed by another repressor, and have both the repressor's mRNA and protein degrade to zero molecules. Asymmetries in the number of operators and the repressor-operator affinity will cause asymmetric pulse widths. In Fig. 7, the model depicted at the right is a symmetric model: repressor-operator affinities at both operator sites on all three operons are 10^{10} M^{-1}. In the model at the left, the repressor-operator affinities of both operator sites on one operon were increased to 10^{12} M^{-1}. Only one marker protein is shown, as the other two were unaffected by the change.

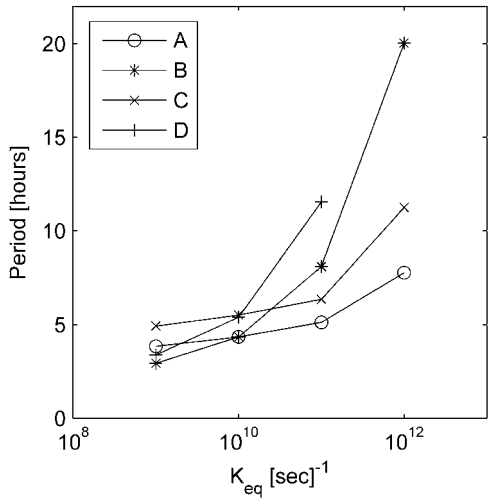

FIGURE 6 Plot of period of oscillation versus repressor-operator affinity for models with (*A* and *B*) two operators per genes and (*C* and *D*) three operators per genes. (*A* and *C*) The affinities of only one set of operators are modified, whereas the other two sets of operators are held at an affinity of 10^{10} M^{-1}. (*B* and *D*) The affinities of all operator sites in all genes are symmetrically altered. The forward kinetic constant is always 10^8 $M^{-1}s^{-1}$.

Variations in the forward rate constant were also briefly investigated. In the preceding trials, forward rate had been held constant at 10^8 $M^{-1}s^{-1}$. Although some repressors, including the *lac* one, are known to exhibit forward binding rates $>10^{10}$ $M^{-1}s^{-1}$ (30), which is greatly in excess of the diffusion limit, such systems are rare. Several models are constructed with the more conservative value of 10^6 $M^{-1}s^{-1}$ and are investigated over the same range of affinities and symmetries as the previous models. In general, these models do not oscillate well, even when the degradation rate of the repressor-operator complex is adjusted to give the same range of affinities. When oscillations do occur, they tend to have longer periods than systems with the same affinity but higher forward rate. For example, a three operator-per-gene symmetric model with an operator-repressor affinity of 10^9 M^{-1} gives a period of 3.39 h and 5.89 h with forward rates of 10^8 $M^{-1}s^{-1}$ and 10^6 $M^{-1}s^{-1}$, respectively. Therefore, the rate of forward binding of a repressor to its operator is shown to be as important as its affinity to the operator.

Other parameters of interest are also investigated. The concentrations of free, functional RNA polymerase and ribosome are difficult to experimentally control. Gene network designs should therefore be robust to variations in these values. The concentration of RNA polymerase in *Escherichia coli* is known to vary with bacterial doubling time. This quantity is not accessible to direct experimentation; however, models indicate values on the order of 100 to 1000 molecules per cell (31). Within this range, five models are evaluated. An initial concentration of RNA polymerases of 100 molecules per cell yields a period of oscillation of 5.66 h, whereas 1000 molecules per cell decreases the period to 5.33 h, which is a difference of only 6%. Variation in initial available ribosome concentration has a somewhat greater effect. With 300 ribosomes per cell, the model oscillates with a period of 4.95 h. When this number is increased to 1000 molecules per cell, the period increases by 19% to 5.89 h. Neither of these quantities leads to a loss of oscillation within the range of values tested. Interestingly, whereas increasing the number of RNA polymerase molecules in a cell leads to a decrease in period, increasing the number of ribosomes has the opposite effect, which is shown in Fig. 8.

The effect of protein degradation rate may also play a role in oscillator design. When the protein products of all three genes are assigned half-lives of 10 min, the period of oscillation is 3.57 h. When the half-lives are increased to 60 min, the period increases to 9.08 h. The same trend is observed when the half-lives of the products of two operators are held constant at 20 min and the half-lives of all of the protein products of the third operator are varied. In this asymmetric case, a half-life of 10 min leads to a period of 4.88 h (versus 3.57 h in the symmetric case) and a half-life of 60 min leads to a period of 6.39 h (versus 9.08 h in the symmetric case). Variation of protein half-life does not lead to a loss of oscillation within the range of values tested.

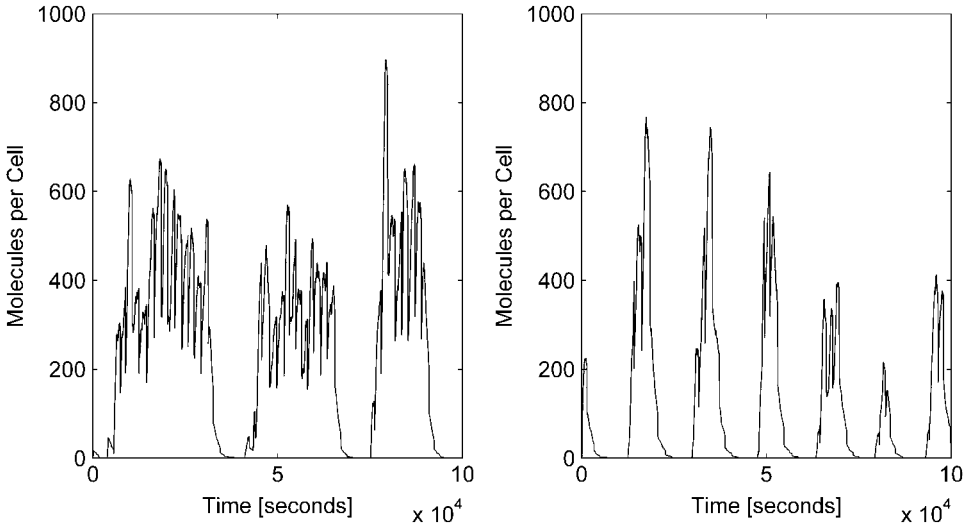

FIGURE 7 Repressor-operator affinity asymmetry causes asymmetric pulse widths. (*Left*) The dynamics of the Dlac protein from a model with two operators per gene, with one operator having a repressor-operator affinity of 10^{12} M^{-1} and the others with repressor-operator affinities of 10^{10} M^{-1}. (*Right*) The dynamics of the Dlac protein from a model with two operators per gene, each having repressor-operator affinities of 10^{10} M^{-1}.

Finally, the effect of mRNA half-life on system oscillation is investigated. As mRNA half-life is varied from 5 min to 15 min for all three mRNA species, the period of oscillations increases from 5.30 to 6.56 h. Fixing the half-lives of two species at 5 min while varying the half-life of the third over the same range yields the same trend, but to a lesser degree. As the third half-life is varied from 5 min to 15 min, the period increases from 5.36 to 5.85 h. Fig. 9 summarizes the data.

CONCLUSIONS

The modified *lac-tet-ara* system described here demonstrates that a gene network created from molecular parts not normally associated with each other nor naturally oscillatory can lead to the expression of a robustly periodic protein product. More interestingly, with the help of inexpensive simulations, a simple set of design rules is created that can guide a first cycle of experiments. The design rules compact the results of these simulations into a few concise statements, which may then be used to more easily construct robust oscillating gene networks.

All of the investigated design parameters affect the period of oscillation. First, increasing the repressor-operator affinity results in a longer period of oscillation. Increasing the number of operators regulating each gene also increases the period of

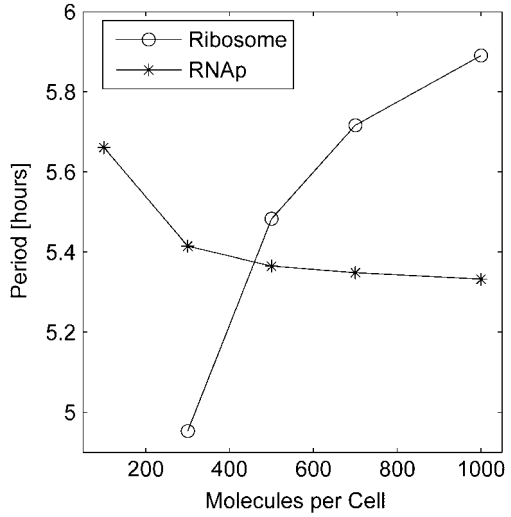

FIGURE 8 Effect of initial ribosome and RNA polymerase numbers on the period of oscillation. The number of initial RNA polymerases and ribosomes is, respectively, varied from 100 to 1000 and 300 to 1000 molecules.

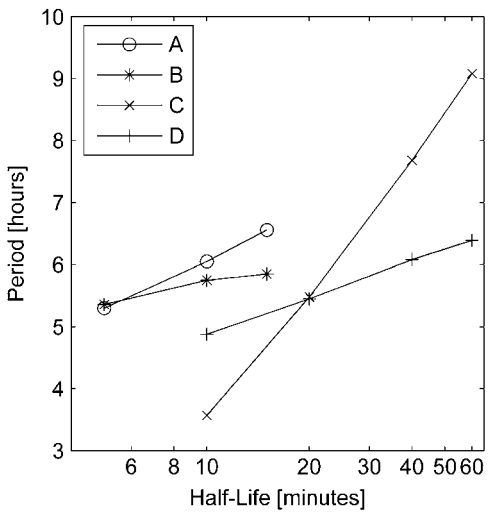

FIGURE 9 Effects of protein and mRNA half-lives on the period of oscillation. (*A*) The half-lives of all mRNA species are symmetrically varied from 5 to 15 min. (*B*) The half-lives of two mRNA species are fixed at 5 min, whereas the half-life of the third species is varied from 5 to 15 min. (*C*) The half-lives of all protein species are symmetrically varied from 10 to 60 min. (*D*) The half-lives of all protein products of two genes are kept constant at 20 min, whereas the half-lives of all protein products of the third gene are varied from 10 to 60 min.

oscillation and makes the repression of each gene more sensitive to the number of repressors, which is abstractly referred to as cooperativity. However, if the operators of the three genes differ in their affinities by larger than two orders of magnitude, then sustained oscillations are not possible. To obtain a desired period of oscillation, one may increase the number of operators and decrease the repressor-operator affinity or vice versa. However, using only one operator generally results in unsustained oscillations, and inserting more than three promoter-overlapping operators is difficult. Second, increasing the half-life of the mRNA or protein of any repressor will increase the period of oscillation. Finally, increasing the number of available RNA polymerases or ribosomes respectively decreases and increases the period of oscillation. Because the RNA polymerases directly compete with repressors for binding to promoter sites, the increase in the number of RNA polymerases has a similar effect as decreasing the repressor-operator affinity. However, because only transcriptional regulation is utilized, increasing the number of ribosomes has a similar effect as increasing the half-lives of the repressors themselves. Although experimentally altering the expression of RNA polymerase and ribosome is not practical, it is important to know how the period of oscillations will change as a result of global shifts in the metabolism of the cell. For example, the period of oscillation is predicted to change when switching from the exponential to stationary growth phases, due only to changes in RNA polymerase and ribosome numbers.

Constructing and testing the variant models described here would be enormously costly. Instead, we use stochastic simulations of a detailed mechanistic model. Although the simulations are subject to multiple assumptions, stochastic modeling of the known interactions should provide a set of verifiable and falsifiable rules. Using the results of a first cycle of targeted experiments, the model may be directly refined to correct invalid assumptions or kinetic parameters. Because the model uses a detailed, mechanistic system of reactions, it is much easier to modify the kinetic characteristics of a particular molecular interaction, which may be directly measured from experiments. Alternative models, which include the extensive use of course-grained or lumped interactions, are more difficult to modify according to new experimental data because they group together multiple biological processes whose independent actions are not fully accounted for.

In the near future, toolboxes will be created of known DNA sequences and protein molecules that exhibit a wide spectrum of well-characterized kinetic parameters. The first successful attempts are already being reported (32). The combination of these molecular components into a synthetic gene network will create novel and useful functions. As these networks become more complex, our ability to intuitively predict their behavior will fail. By using detailed mechanistic models and stochastic simulation techniques, one can more quickly determine the necessary molecular components and network connectivities that produce a desired dynamical behavior.

This work was supported by grants from the National Science Foundation (BES-0425882 and EEC-0234112). Computational support from the Minnesota Supercomputing Institute is gratefully acknowledged. This work was also supported by the National Computational Science Alliance under TG-MCA04N033.

REFERENCES

1. Elowitz, M. B., and S. Leibler. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*. 403:335–338.
2. Kaern, M., W. J. Blake, and J. J. Collins. 2003. The engineering of gene regulatory networks. *Annu. Rev. Biomed. Eng*. 5:179–206.
3. Goldbeter, A. 2002. Computational approaches to cellular rhythms. *Nature*. 420:238–245.
4. Stelling, J., E. D. Gilles, and F. J. Doyle 3rd. 2004. Robustness properties of circadian clock architectures. *Proc. Natl. Acad. Sci. USA*. 101:13210–13215.
5. Youan, B. B. 2004. Chronopharmaceutics: gimmick or clinically relevant approach to drug delivery? *J. Controlled Release*. 98:337–353.
6. Smolen, P., D. A. Baxter, and J. H. Byrne. 2002. A reduced model clarifies the role of feedback loops and time delays in the Drosophila circadian oscillator. *Biophys. J*. 83:2349–2359.
7. Vilar, J. M., H. Y. Kueh, N. Barkai, and S. Leibler. 2002. Mechanisms of noise-resistance in genetic oscillators. *Proc. Natl. Acad. Sci. USA*. 99:5988–5992.
8. Barkai, N., and S. Leibler. 2000. Circadian clocks limited by noise. *Nature*. 403:267–268.
9. Hasty, J., M. Dolnik, V. Rottschafer, and J. J. Collins. 2002. Synthetic gene network for entraining and amplifying cellular oscillations. *Phys. Rev. Lett*. 88:148101.
10. Elowitz, M. B., A. J. Levine, E. D. Siggia, and P. S. Swain. 2002. Stochastic gene expression in a single cell. *Science*. 297:1183–1186.
11. Gilbert, W., and B. Muller-Hill. 1970. The Lactose Operon. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
12. Vogel, U., and K. F. Jensen. 1994. The RNA chain elongation rate in Escherichia coli depends on growth rate. *J. Bacteriol*. 176:2807–2813.
13. Sorenson, M. A., and S. Pedersen. 1991. Absolute in vivo translation rates of individual codons in Escherichia coli. *J. Mol. Biol*. 222:265–280.
14. Kedracka-Krok, S., and Z. Wasylewski. 1999. Kinetics and equilibrium studies of Tet repressor-operator interaction. *J. Protein Chem.* 18: 117–125.
15. Bertrand-Burggraf, E., J. F. Lefevre, and M. Daune. 1984. A new experimental approach for studying the association between RNA polymerase and the tet promoter of pBR322. *Nucleic Acids Res*. 12: 1697–1706.
16. Stickle, D. F., K. M. Vossen, D. A. Riley, and M. G. Fried. 1994. Free DNA concentration in E. coli estimated by an analysis of competition for DNA binding proteins. *J. Theor. Biol*. 168:1–12.
17. Zhang, X., T. Reeder, and R. Schleif. 1996. Transcription activation parameters at ara pBAD. *J. Mol. Biol*. 258:14–24.
18. Frank, D. E., R. M. Saecker, J. P. Bond, M. W. Capp, O. V. Tsodikov, S. E. Melcher, M. M. Levandoski, and M. T. J. Record. 1997. Thermodynamics of the interactions of Lac repressor with variants of the symmetric Lac operator: effects of converting a consensus site to a non-specific site. *J. Mol. Biol*. 267:1186–1206.
19. Wissmann, A., R. Baumeister, G. Muller, B. Hecht, V. Helbl, K. Pfleiderer, and W. Hillen. 1991. Amino acids determining operator binding specificity in the helix-turn-helix motif of Tn10 Tet repressor. *EMBO J*. 10:4145–4152.
20. Helbl, V., B. Tiebel, and W. Hillen. 1998. Stepwise selection of TetR variants recognizing tet operator 6C with high affinity and specificity. *J. Mol. Biol*. 276:319–324.

21. Helbl, V., and W. Hillen. 1998. Stepwise selection of TetR variants recognizing tet operator 4C with high affinity and specificity. *J. Mol. Biol.* 276:313–318.

22. Gillespie, D. T. 1976. A general method for numerically simulating the stochastic time evolution of coupled chemical reactions. *J. Comput. Phys.* 22:403–434.

23. Gibson, M. A., and J. Bruck. 2000. Efficient exact stochastic simulation of chemical systems with many species and many channels. *J. Phys. Chem. A.* 104:1876–1889.

24. Cao, Y., H. Li, and L. Petzold. 2004. Efficient formulation of the stochastic simulation algorithm for chemically reacting systems. *J. Chem. Phys.* 121:4059–4067.

25. Gillespie, D. T. 2000. The chemical Langevin equation. *J. Chem. Phys.* 113:297–306.

26. Salis, H., and Y. Kaznessis. 2005. Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions. *J. Chem. Phys.* 122:054103.

27. Giannakis, G. B. 1998. Cyclostationary signal analysis. *In* Digital Signal Processing Handbook. V. K. Madisetti and D. Williams, editors. CRC Press, Boca Raton FL. 17-1–17.31.

28. Lutz, R., and H. Bujard. 1997. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1–I2 regulatory elements. *Nucleic Acids Res.* 25:1203–1210.

29. Lutz, R., T. Lozinski, T. Ellinger, and H. Bujard. 2001. Dissecting the functional program of Escherichia coli promoters: the combined mode of action of Lac repressor and AraC activator. *Nucleic Acids Res.* 29: 3873–3881.

30. Halford, S. E., and J. F. Marko. 2004. How do site-specific DNA-binding proteins find their targets? *Nucleic Acids Res.* 32:3040–3052.

31. Bremer, H., P. Dennis, and M. Ehrenberg. 2003. Free RNA polymerase and modeling global transcription in Escherichia coli. *Biochimie.* 85: 597–609.

32. Endy, D., and R. Brent. 2001. Modelling cellular behaviour. *Nature.* 409:391–395.