Optimizing Genetic Circuits by Global Sensitivity Analysis

Xiao-jiang Feng,* Sara Hooshangi,[†] David Chen,[†] Genyuan Li,* Ron Weiss,[†‡] and Herschel Rabitz*

*Department of Chemistry, [†]Department of Electrical Engineering, and [‡]Department of Molecular Biology,
Princeton University, Princeton, New Jersey

ABSTRACT   Artificial genetic circuits are becoming important tools for controlling cellular behavior and studying molecular biosystems. To genetically optimize the properties of complex circuits in a practically feasible fashion, it is necessary to identify the best genes and/or their regulatory components as mutation targets to avoid the mutation experiments being wasted on ineffective regions, but this goal is generally not achievable by current methods. The Random Sampling—High Dimensional Model Representation (RS-HDMR) algorithm is employed in this work as a global sensitivity analysis technique to estimate the sensitivities of the circuit properties with respect to the circuit model parameters, such as rate constants, without knowing the precise parameter values. The sensitivity information can then guide the selection of the optimal mutation targets and thereby reduce the laboratory effort. As a proof of principle, the in vivo effects of 16 pairwise mutations on the properties of a genetic inverter were compared against the RS-HDMR predictions, and the algorithm not only showed good consistency with laboratory results but also revealed useful information, such as different optimal mutation targets for optimizing different circuit properties, not available from previous experiments and modeling.

INTRODUCTION

In recent years, an increasing number of studies have focused on constructing simple synthetic genetic circuits with desired properties (Becskei and Serrano, 2000; Elowitz and Leibler, 2000; Gardner et al., 2000; Hasty et al., 2000, 2001a,b, 2002a,b; Weiss, 2001; Basu et al., 2002; Guet et al., 2002; Weiss and Basu, 2002; Yokobayashi et al., 2002; Atkinson et al., 2003; Francois and Hakim, 2004). The importance of these studies is twofold. First, the synthetic systems can serve as a basis to understand the workings of natural systems, which are usually much more complicated and difficult to unravel. Second, artificial genetic circuits may also act as tiny ''programs'' to control cellular behavior in specified manners, thus providing various potential applications in biotechnology, medicine, environmental science, and other areas. Although the research in several prototype circuits has provided insights and offered great promise in fulfilling both targets, designing and engineering in vivo genetic circuits remains a difficult task. One common obstacle is to design and optimize individual circuit components (or ''devices'') and their interconnections so that they can be integrated to form functional circuits. The consequence of linking unmatched devices was clearly illustrated in constructing a simple genetic inverter (Weiss, 2001; Weiss and Basu, 2002), where the output of a circuit component A represses another component B, realizing the inverter function. When the two components do not match (e.g., if the ''low'' output of A is still high enough to repress B, then the output of B will always stay at the ''low'' state despite the level of A), the result can be a nonfunctional genetic inverter.

Three general strategies have been employed for designing and optimizing genetic circuits. The first method is to rationally design the circuits, exemplified by Weiss and co-workers in building a genetic inverter, where gene mutations that transformed the circuit into a functional unit were guided by modeling and simulations (Weiss, 2001; Weiss and Basu, 2002). The second approach was described by the work of Guet et al. (2002) in which they randomly reshuffled the connectivity of a simple genetic network and obtained multiple circuits with different logic functions. The third approach was to employ directed evolution techniques to optimize circuit properties, either in vivo (Yokobayashi et al., 2002) or in silico (Francois and Hakim, 2004).

In principle, all three strategies above, especially the first and the third approaches, can be employed in building more complex genetic circuits with specified properties. However, as the complexity of the circuits rises, it becomes increasingly difficult to apply these strategies. For example, as circuits involve a larger number of molecular species for optimization, it becomes increasingly necessary to predetermine the optimal molecular targets (e.g., genes and their regulatory components) where genetic mutations can most effectively achieve the specified circuit properties. Without suitable guidance the random mutations or directed evolution experiments can be wasted on ineffective regions, thus becoming costly or even practically prohibitive. However, the rational design method above usually cannot identify the optimal mutation targets, because 1), the circuits are highly nonlinear networks with complex behavior generally not amenable to traditional modeling and analysis techniques and 2), although the basic structures of the circuits (hence the models) are well defined, the model parameters (e.g., the kinetic rate constants) usually contain significant uncertainties, which render any deterministic analysis methods inadequate.

In this work, a global, nonlinear, stochastic sensitivity analysis technique called the Random Sampling—High Dimensional Model Representation (RS-HDMR) algorithm (Li et al., 2001, 2002; Wang et al., 2003) is employed to assist in overcoming the above difficulties. The RS-HDMR algorithm can provide reliable pre-experimental estimates on the sensitivities of the circuit properties (e.g., the inverter gain g) with respect to broad scale variations in the model parameters (e.g., the translation rate constant k_{trl} associated with a certain protein) without knowing their precise values. The sensitivity information can then be used to guide the selection of mutation targets. For example, if g is highly sensitive to variations in k_{trl}, it suggests that genetic mutations that change the k_{trl} value can be effective in optimizing the inverter gain. Conversely, if the sensitivity of g to k_{trl} is low, it implies that the corresponding mutations should be avoided.

As a proof of principle, we examined 16 pairwise mutations on a well-studied genetic inverter and compared the in vivo effects of the mutations on the circuit properties with the RS-HDMR predictions. The theoretical results not only showed satisfactory consistency with the laboratory observations, but also provided useful insights unavailable from previous modeling and experimental studies. For example, the analysis revealed quantitatively that the inverter output (enhanced yellow fluorescent protein steady-state concentration) is more sensitive to mutations in the ribosome-binding site (RBS) upstream of the cI coding region than mutations in the O_R1 region of the P_R promoter. The analysis also showed that these mutations have larger effects on the enhanced yellow fluorescent protein (EYFP) concentration at high input (IPTG concentration) levels than at low IPTG levels, whereas the effects of EYFP transcription and translation on EYFP concentrations are fairly stable against variations in IPTG levels. At last, RS-HDMR clearly identified that mutations affecting the transcription and translation of EYFP serve the best for adjusting EYFP concentrations at different IPTG levels, whereas the RBS mutations are the most effective in optimizing the gain and the slope of the inverter, all of which are not intuitively evident or predictable from simple analyses. The successful application of RS-HDMR in this work establishes the capability of the algorithm for accelerating the optimization of complex genetic circuits and potentially enhancing the utility of genetic engineering. More importantly, similar principles and algorithms can be applied in the mechanistic studies of naturally occurring bionetworks, as discussed at the end of the article.

MATERIALS AND METHODS

Plasmids

Plasmids pINV-110, pINV-112-R1, pINV-112-R2, and pINV-112-R3 (simplified as p110, pR1, pR2, and pR3, respectively) encode λ-repressor CI and enhanced cyan fluorescent protein under the control of lac promoter P_{lac} (Fig. 1). The four plasmids differ in the RBS sequence located upstream of the cI coding region with translation efficiency p110 > pR1 > pR2 > pR3 (Weiss, 2001; Weiss and Basu, 2002). They contain a kanamycin resistance marker and the p15A replication origin. Plasmids pINV-107, pINV-107-MUT4, pINV-107-MUT5, and pINV-107-MUT6 (denoted as p107, pM4, pM5, and pM6, respectively) encode enhanced yellow fluorescent protein (EYFP) under the control of the synthetic λ right promoter (λP_{RO12}, a partial $\lambda P_{(R)}$ with only O_R1 and O_R2). They differ in the operator binding sequence O_R1 with repressor/operator binding affinity p107 > pM4 > pM5 $\approx$ pM6 (Weiss, 2001; Weiss and Basu, 2002). These plasmids contain an ampicillin resistance marker and the ColE1 replication origin.

Circuit performance measurements

Measurements of the circuit properties were carried out in liquid culture using a fluorescence-activated cell sorter. The EYFP fluorescence level was the ultimate output of the circuit, and enhanced cyan fluorescent protein was not used. *Escherichia coli* cells harboring the plasmids were first grown overnight to stationary phase in LB medium containing appropriate antibiotics. The cultures were then diluted 500-fold into 2 ml of fresh LB medium containing varying amounts of isopropyl-β-D-thiogalactoside (IPTG) and the same antibiotics. The cells were then grown for 6 h at 37°C to log phase (OD $\approx$ 0.2), harvested, centrifuged, washed, and suspended in 0.5 ml of PBS (0.22-μm filter sterilized, pH 7.5). All cell samples were cultured in triplicate and their EYFP fluorescence levels were measured (Epics Altra flow cytometer, Beckman Coulter, Fullerton, CA) and calibrated using *SPHERO* calibration particles (RCP-30-5A, Spherotech, Libertyville, IL). The calibrated fluorescence levels were reported in a molecules-of-equivalent fluorescein (MEFL) unit, which is proportional to the EYFP concentration. The mean fluorescence values of the triplicates were used for analysis and the relative errors were $\sim\pm10\%$.

FIGURE 1 The plasmid diagram of the genetic inverter, the logic diagram (*upper right*), and the truth table (*lower right*). Plasmid pINV-110 constitutively expresses the LacI repressor, which always stays at a high level in the experiments. The P_{lac} promoter serves as the IMPLIES gate, with the inputs being the repressor LacI and the externally added regulator IPTG, and the output being CI. CI is the input of the inverter λP_{RO12} and represses the synthesis of the fluorescent output EYFP. ''1'' and ''0'' in the truth table represent high and low states of the circuit components, respectively.

The mechanistic model

The components of this genetic inverter have been intensively studied (Miller and Reznikoff, 1980; Lewin, 2000), providing the basis for the chemical kinetics model. The model of the genetic inverter used in this work (Fig. 2) contains 13 chemical species and 18 rate constants. Eq. F1 describes the repression of the P_{lac} promoter by the LacI tetramer. Eq. F2 represents the binding of LacI by IPTG. An average of two IPTG molecules are assumed to bind effectively to one tetramer (Miller and Reznikoff, 1980; Yildirim and Mackey, 2003). Eqs. F3 and F4 represent transcription from P_{lac} to produce mRNA for CI (mRNA(CI)) and the decay of mRNA(CI). Eqs. F5 and F6 describe the synthesis of CI from mRNA(CI) and the decay of CI. In Eqs. F3 and F5, the RNA polymerase and the ribosome RNA are assumed to be in excess, hence they do not appear in the equations. Eqs. F7–F9 describe the dimerization of CI and the cooperative repression of the λP_{RO12} promoter by the dimer. Eqs. F10–F13 represent transcription and translation of EYFP starting from λP_{RO12}, as well as the decay of the mRNA and EYFP. Again, the RNA polymerase and the ribosome RNA are assumed to be in excess. All the rate constant values and initial concentrations were either obtained or derived from relevant sources (Miller and Reznikoff, 1980; Lewin, 2000; Yildirim and Mackey, 2003; Weiss, 2001).

The genetic circuit

Fig. 1 shows the implementation and the logic circuit diagram of the genetic circuit. It involves two separate plasmids, one including the LacI repressor/P_{lac} promoter, and the other coding for the CI repressor/λP_{RO12} promoter. In the first plasmid, LacI protein is expressed from the P_{lacIq} promoter and represses the P_{lac} promoter, which otherwise initiates the expression of CI protein. CI binds the operator binding sites of the λP_{RO12} promoter on the second plasmid, repressing the synthesis of EYFP.

The behavior of the circuit can be controlled by adding an inducer, IPTG. LacI and IPTG form a complex with reduced binding affinity to the P_{lac} promoter, thus the presence of IPTG helps to increase the production of CI, which in turn reduces the synthesis of EYFP. Since LacI always stays at a relatively high level, the circuit can be regarded as a genetic inverter: when the input IPTG level is high, the output EYFP concentration is low; and when IPTG concentration decreases, EYFP production increases. Due to the cooperative binding of CI to λP_{RO12} (Miller and Reznikoff, 1980), the steady-state transfer curve (EYFP versus IPTG curve) of the inverter is normally an inverse sigmoidal (see Fig. 3 for examples), and the characteristics of the curve (e.g., the gain) are determined by the reaction rate constant values.

For the circuit to function correctly, its two components must match each other. Biochemically, this means that a ''high'' output (i.e., CI concentration) from the first component must be sufficient to repress most of λP_{RO12} (the input of the second component), and a ''low'' CI concentration should have little effect on λP_{RO12} expression. Two mismatched genetic components can generate nonfunctional circuits. For example, in the initial attempt to build the inverter (also the circuit p110:p107 in this study), the ''low'' output of the first component was interpreted as ''high'' input by the second component, because the binding affinity of λP_{RO12} with CI was so large that a low concentration of CI was sufficient to repress λP_{RO12}. Consequently, EYFP always stayed at very low levels, and the transfer curve was essentially flat (Weiss, 2001; Weiss and Basu, 2002).

The rational optimization of this circuit has been described (Weiss, 2001; Weiss and Basu, 2002). An ordinary differential equation model was first constructed to simulate the dynamic and steady-state behavior of the circuit. Then a few rate constants in the model were changed and their effects on the transfer curve were recorded. Since the rate constant variations correspond directly to mutations on one or more genes or regulatory sites, the simulations provided guidance as to where and how (increased or reduced

$$LacI + P_{lac} \underset{k_1}{\overset{k_0}{\rightleftharpoons}} LacI.P_{lac} \qquad (F1)$$

$$LacI + 2IPTG \underset{k_3}{\overset{k_2}{\rightleftharpoons}} LacI.IPTG_2 \qquad (F2)$$

IPTG controlled LacI repression of P_{lac}

$$P_{lac} \xrightarrow{k_4} P_{lac} + mRNA(CI) \qquad (F3)$$

$$mRNA(CI) \xrightarrow[decay]{k_5} \qquad (F4)$$

transcription and decay of mRNA(CI)

$$mRNA(CI) \xrightarrow{k_6} mRNA(CI) + CI \qquad (F5)$$

$$CI \xrightarrow[decay]{k_7} \qquad (F6)$$

translation and decay of CI protein

$$2CI \underset{k_9}{\overset{k_8}{\rightleftharpoons}} CI_2 \qquad (F7)$$

$$P_R + CI_2 \underset{k_{11}}{\overset{k_{10}}{\rightleftharpoons}} P_R.CI_2 \qquad (F8)$$

$$P_R.CI_2 + CI_2 \underset{k_{13}}{\overset{k_{12}}{\rightleftharpoons}} P_R.CI_4 \qquad (F9)$$

repression of P_R by CI

$$P_R \xrightarrow{k_{14}} P_R + mRNA(EYFP) \qquad (F10)$$

$$mRNA(EYFP) \xrightarrow[decay]{k_{15}} \qquad (F11)$$

transcription and decay of mRNA(EYFP)

$$mRNA(EYFP) \xrightarrow{k_{16}} mRNA(EYFP) + EYFP \qquad (F12)$$

$$EYFP \xrightarrow[decay]{k_{17}} \qquad (F13)$$

translation and decay of EYFP protein

FIGURE 2 The reaction mechanism for the genetic inverter in Fig. 1. The nominal values of the rate constants are: $k_0 = 7.0 \times 10^3 \ \mu M^{-1} \ s^{-1}$, $k_1 = 6.0 \times 10^{-4} \ s^{-1}$, $k_2 = 1.0 \times 10^{-3} \ \mu M^{-1} \ s^{-1}$, $k_3 = 1.0 \times 10^{-2} \ s^{-1}$, $k_4 = 1.5 \times 10^{-2} \ s^{-1}$, $k_5 = 4.0 \times 10^{-3} \ s^{-1}$, $k_6 = 3.0 \times 10^{-2} \ s^{-1}$, $k_7 = 5.0 \times 10^{-3} \ s^{-1}$, $k_8 = 0.1 \ \mu M^{-1} \ s^{-1}$, $k_9 = 2.0 \times 10^{-3} \ s^{-1}$, $k_{10} = 1.0 \ \mu M^{-1} \ s^{-1}$, $k_{11} = 2.0 \times 10^{-3} \ s^{-1}$, $k_{12} = 5.0 \ \mu M^{-1} \ s^{-1}$, $k_{13} = 2.0 \times 10^{-3} \ s^{-1}$, $k_{14} = 1.5 \times 10^{-2} \ s^{-1}$, $k_{15} = 4.0 \times 10^{-3} \ s^{-1}$, $k_{16} = 3.0 \times 10^{-2} \ s^{-1}$, and $k_{17} = 5.0 \times 10^{-3} \ s^{-1}$. The initial concentrations are 0.02 μM for the LacI tetramer, P_{lac}, and λP_{RO12} (denoted as P_R), 0–$10^3 \ \mu M$ for IPTG, and zero for all other species.
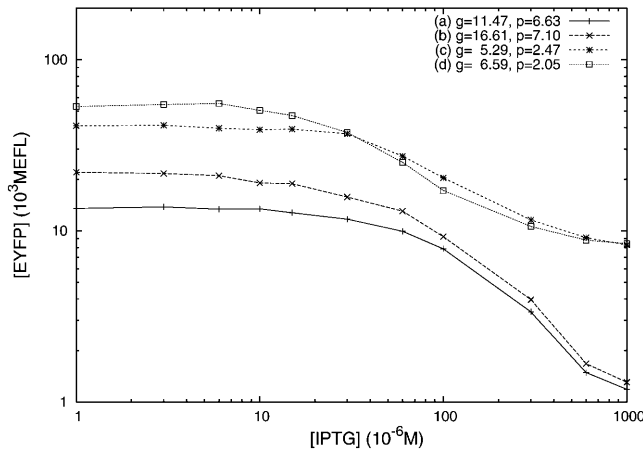
FIGURE 3 The steady-state transfer curves of four genetic inverters (measured at 12 IPTG levels) and their gain g and slope p. (a) pR1:pM4, (b) pR2:pM4, (c) pR1:pM5, and (d) pR2:pM5. The value g is calculated by Eq. 6, and p is determined by Eq. 8.

activity) the mutations should be performed to achieve desired circuit behavior. Following the simulation results, the translation rate of the CI protein was first reduced by site-directed mutagenesis to the RBS upstream of cI coding region. Gene mutations were then performed at the O_R1 operator binding site of the λP_{RO12} promoter to weaken repressor (CI) binding. The experiments yielded circuits with enhanced performance qualitatively consistent with the simulation results. In a subsequent article, Yokobayashi et al. (2002) applied directed evolution techniques to mutate the original, nonfunctional circuit on both the cI gene and the RBS and generated functional circuits. In both studies, the authors picked the molecular candidates for mutation experiments based on their insight into the system as well as simple modeling and analysis. This approach, however, could not determine if the selected mutation targets would be the most effective ones in enhancing the circuit performance, and this problem can become serious in engineering more complex circuits where the laboratory costs may be very high if the experiments are not carried out on the optimal mutation targets.

The RS-HDMR algorithm

Given a quantitative model of a genetic circuit, the RS-HDMR algorithm can provide reliable estimates for the sensitivities of the circuit properties with respect to the model parameters without knowing their precise values. In this fashion, RS-HDMR can estimate the optimal mutation targets based on the sensitivity values. Both deterministic and stochastic models have been used to understand the behavior of genetic circuits (McAdams and Arkin, 1998; Bower, 2001; Hasty et al., 2001c; Jong, 2002). In this work, RS-HDMR is applied using deterministic models involving ordinary differential equations (ODEs), although the principles and techniques can be employed for other types of models (e.g., stochastic or spatiotemporal, etc.). Consider a system containing N species $\mathbf{x} = (x_1, x_2, \ldots, x_N)$ and M reaction rate constants $\mathbf{k} = (k_1, k_2, \ldots, k_M)$ with its dynamic behavior described by the ODEs,

$$\frac{dX_n}{dt} = f_n(\mathbf{X}, \mathbf{k}) \quad n = 1, 2, \ldots, N, \qquad (1)$$

where X_n is the concentration of x_n. In this model, gene mutations correspond to changing one or more of the rate constants k_m, and the properties of the genetic circuit are represented by the X_n values (temporal or steady-state) or functions of them. The aim of RS-HDMR analysis is to quantitatively identify the effects of the variations in $\mathbf{k}$ on the circuit

properties. RS-HDMR is a global sensitivity analysis technique that can decompose the high-dimensional, nonlinear contributions of $\mathbf{k}$ to the network properties (represented by their total sensitivity) into a hierarchy of low-dimensional terms. Calculations by RS-HDMR require only the ODE model, an estimate of the initial conditions X_n^0 (to be used for ODE integration), and an estimate of the dynamic range $[k_m^<, k_m^>]$ to ''explore'' for each rate constant k_m. Since the genetic circuit components are built from well-studied natural networks, all these requirements are usually satisfied, or can be satisfied by performing a few additional experiments.

To estimate the total sensitivity of X_n to $\mathbf{k}$, for example, normally several thousand sets of randomly chosen rate constants $\mathbf{k}^s (s = 1, 2, \ldots, S)$ are generated over the estimated ranges $[\mathbf{k}^<, \mathbf{k}^>]$. The transient concentration profile of X_n is calculated for each $\mathbf{k}^s$ by integrating the ODEs, and the total sensitivity $\sigma_t(X_n)$ of X_n at time t is calculated as a relative standard deviation

$$\sigma_t(X_n) = \left[\frac{1}{S} \sum_{s=1}^{S} (X_{n,t}^s)^2 - \left(\frac{1}{S} \sum_{s=1}^{S} (X_{n,t}^s) \right)^2 \right]^{1/2} \bigg/ w_{n,t}, \qquad (2)$$

where $X_{n,t}^s$ is the concentration of x_n at time t for sample s, and $w_{n,t}$ is a weight factor that normalizes the absolute standard deviation of $X_{n,t}$. Similarly, the total sensitivity $\sigma^*(X_n)$ at steady state can be calculated by

$$\sigma^*(X_n) = \left[\frac{1}{S} \sum_{s=1}^{S} (X_n^{*,s})^2 - \left(\frac{1}{S} \sum_{s=1}^{S} (X_n^{*,s}) \right)^2 \right]^{1/2} \bigg/ w_n, \qquad (3)$$

where $X_n^{*,s}$ is the steady-state concentration of x_n for sample s, and w_n is a weight factor. The total sensitivity $\sigma_t(X_n)$ in Eq. 2 is decomposed into a set of contributions,

$$\sigma_t^2(X_n) = \sum_{m=1}^{M} \sigma_t^2(X_n, k_m)$$
$$+ \sum_{1 \le m < m' \le M} \sigma_t^2(X_n, (k_m, k_{m'})) + \ldots, \qquad (4)$$

where the first-order term $\sigma_t(X_n, k_m)$ represents the effect that the single independent variable k_m has on X_n, and the second-order term $\sigma_t(X_n, (k_m, k_{m'}))$ reflects the cooperative influence of k_m and $k_{m'}$ on X_n, etc. The steady-state total sensitivity $\sigma^*(X_n)$ similarly can be decomposed by

$$\sigma^{*2}(X_n) = \sum_{m=1}^{M} \sigma^2(X_n^*, k_m)$$
$$+ \sum_{1 \le m < m' \le M} \sigma^2(X_n^*, (k_m, k_{m'})) + \ldots. \qquad (5)$$

The details of the decomposition are discussed elsewhere (Li et al., 2001, 2002; Wang et al., 2003). In the same fashion, the sensitivities of any circuit property $y = y(\mathbf{X}, t)$ to $\mathbf{k}$ can be calculated by replacing X_n in the above equations by y, as long as $y(\mathbf{X}, t)$ can be determined analytically or numerically.

In principle, other sensitivity analysis methods can also be employed (Saltelli et al., 2000), but we choose RS-HDMR for three reasons. First, it is a global analysis technique that works well even when the input variables (rate constants in this case) contain significant uncertainties. This characteristic is especially important for its applications in biology where model parameters usually cannot be identified with high precision. Second, it is a nonlinear algorithm, and the decomposed sensitivities (even the first-order terms) are generally also nonlinear, making it a suitable algorithm for analyzing bionetworks. Finally, the first-, second-, and higher-order sensitivity terms are physically meaningful (representing the independent, pairwise, and higher-order contributions of the input variables to specified

circuit properties), and can provide clear guidance for determining the independent and cooperative molecular targets for gene mutations.

RESULTS AND DISCUSSIONS

Circuit synthesis and performance measurements

To test the reliability of the RS-HDMR algorithm in identifying the mutation targets, we synthesized 16 pairwise mutants of this genetic inverter, each of which combines one mutant of the first plasmid (p110, pR1, pR2, and pR3) with one mutant of the second plasmid (p107, pM4, pM5, and pM6). The steady-state EYFP fluorescence for each of the 16 plasmids was then measured at [IPTG] = 0 and [IPTG] = 1 mM in triplicate. The calibrated MEFL values are shown in Table 1, where the columns correspond to RBS mutations and the rows are O_R1 mutations. For all the plasmids (except for p110:p107, which is nonfunctional) the MEFL values are higher with [IPTG] = 0 than with [IPTG] = 1 mM, showing appropriate inverter behavior. At both IPTG concentrations, the EYFP levels are generally higher going to the right and the bottom of Table 1, in agreement with previous studies (Weiss, 2001; Weiss and Basu, 2002), which showed that EYFP levels increase with lower RBS binding affinity and lower O_R1 binding affinity (note that p110 > pR1 > pR2 > pR3 in translation efficiency and p107 > pM4 > pM5 ≈ pM6 in O_{R1} binding affinity). The MEFL levels of four circuits (pR1:pM4, pR1:pM5, pR2:pM4, and pR2:pM5) were also measured at 12 different IPTG concentrations, and their transfer curves are shown in Fig. 3.

Modeling and sensitivity analysis

Based on the model described in Materials and Methods, the sensitivities of EYFP steady-state concentrations at 10 different IPTG levels (between 0 and 5 mM) to variations in the 18 rate constants were calculated using the RS-HDMR algorithm. First, the 18 rate constants were randomly sampled simultaneously from within their corresponding dynamic ranges $[k_m^<, k_m^>]$. Three different dynamic ranges

were used for each rate constant, spanning two, three, and four orders of magnitudes around the nominal values of each rate constant (listed in Fig. 2). The rate constants were transformed to a logarithmic scale to ensure an even distribution over the large space. At each IPTG level, the steady-state concentration of EYFP ([EYFP]*,s) was obtained for each set of random rate constants $\mathbf{k}^s$ by integrating the ODEs until the variation in EYFP concentration was <0.1% within 1000 s. It was assumed that an EYFP concentration of 0.02 μM corresponds to 10 EYFP molecules per cell, and any steady-state EYFP concentration lower than it was removed from the simulations. For each IPTG level, when $S = 10,000$ good samples were obtained, the total sensitivities were calculated using Eq. 3 and normalized by the mean of the EYFP steady-state concentrations (i.e., $w_n = \frac{1}{S}\sum_{s=1}^{S}$[EYFP]*,s in Eq. 3). The first- and second-order sensitivities of [EYFP]*,s with respect to the 18 rate constants were then calculated using Eq. 5 (the detailed procedure is described in Li et al., 2001, 2002 and Wang et al., 2003). Calculations using the three different dynamic rate constant ranges gave similar sensitivity values, and all analyses in this work are based on results obtained from the smallest dynamic ranges. This work is focused on how the circuit properties are affected by rate constant changes, and in the same fashion, RS-HDMR can also calculate sensitivities to variations in other circuit variables, which will be addressed in future research.

Comparison between theoretical and experimental results

In the 16 inverters, genetic mutations were induced on both the RBS of mRNA(CI) and the O_R1 region of the P_R promoter. The former mutations correspond to variations in k_6 (see Fig. 2), and the latter correspond to changes in k_{10}, k_{11}, and possibly k_{12} and k_{13} (indirectly). Fig. 4 plots the total

TABLE 1 Steady-state MEFL levels (unit: 10^3 MEFL) of the 16 genetic inverters at [IPTG] = 0 and [IPTG] = 1 mM

	Steady-state MEFL levels			
[IPTG] = 1 mM	p107	pM4	pM5	pM6
p110	0.15	0.36	0.70	1.00
pR1	0.15	1.13	5.45	6.80
pR2	0.20	1.21	6.10	8.85
pR3	0.20	2.05	6.20	8.90
[IPTG] = 0 mM	p107	pM4	pM5	pM6
p110	0.15	0.65	7.85	15.35
pR1	1.45	12.84	40.95	34.20
pR2	8.00	20.32	41.95	34.10
pR3	7.30	30.54	45.55	36.10

The columns correspond to RBS mutations, and the rows correspond to O_R1 mutations.

FIGURE 4 The total sensitivity of EYFP steady-state concentrations at various IPTG levels, as well as the first-order sensitivities contributed by k_6, k_{10}, k_{11}, k_{12}, k_{13}, and k_{14}.

sensitivities of EYFP concentration at 10 IPTG levels as well as the first-order sensitivities to k_6, k_{10}, k_{11}, k_{12}, and k_{13}. k_{14} to k_{17} in general have the highest first-order sensitivities, and since their values are similar at all IPTG levels, only the k_{14} sensitivities are plotted in Fig. 4. At all 10 IPTG levels, the sum of the first-order terms contributes to >60% of the total sensitivities, and the contribution of the first- and the second-order terms together is >90%. Since all the important second-order terms involve rate constants with important first-order contributions, the first-order sensitivities can serve as semiquantitative indicators of the corresponding rate constants' general influence on EYFP concentrations.

Four conclusions can be drawn from the sensitivity values:

1. The normalized total sensitivity shows little variations against different IPTG levels.
2. At low [IPTG] (<10 μM), the EYFP level is relatively insensitive to variations in k_6, k_{10}, k_{11}, k_{12}, and k_{13}, and their sensitivities are significantly higher at high [IPTG] than at low [IPTG].
3. Among the five rate constants above, k_6 is more effective in influencing EYFP levels than the other four rate constants.
4. k_{14} to k_{17} in general contribute the most to EYFP variations (especially at low IPTG levels), and their first-order sensitivities change very little over different IPTG levels.

Conclusion *1* suggests that, statistically the influence of a large number of random multipoint mutations on the EYFP levels (corresponding to the total sensitivities) is independent of IPTG concentrations, although the reason is unclear. Conclusion *2* serves as the best comparison against the laboratory results in Table 1. For example, the EYFP level increases 8.90/1.00 = 8.9 times from p110:pM6 to pR3:pM6 with [IPTG] = 1 mM, whereas the increase is 36.10/15.35 = 2.35 times with [IPTG] = 0, indicating a more pronounced influence of CI translation efficiency (involving k_6) on EYFP concentrations at higher IPTG levels. This trend is clearly demonstrated in the computational results (Fig. 4), where the corresponding first-order sensitivity of [EYFP] to k_6 is higher at high [IPTG]. Similarly, RS-HDMR predicts that the first-order sensitivities to k_{10}, k_{11}, k_{12}, k_{13} should rise with increasing [IPTG], whereas in Table 1, the EYFP level increases 8.90/2.05 = 4.3 times from pR3:pM4 to pR3:pM6 with [IPTG] = 1 mM and it increases 36.10/30.54 = 1.2 times with [IPTG] = 0. The same comparisons can be made across other columns and rows of Table 1, with the RS-HDMR predictions mostly consistent with laboratory results. There are a few exceptions (e.g., the column containing p107 as well as the double plasmid p110:pM4), but all of them involve very low MEFL values, which may be below the detection sensitivity (background noise level) of fluorescent proteins. It is difficult to compare Conclusion *3* with the experimental results, because the five rate constants change to different extents in the pairwise mutants, and their

quantitative values are not available, thus normalizing (and then comparing) the columns against the rows of Table 1 is not possible.

Conclusion *4* is also biophysically reasonable. The model (Fig. 2) indicates that most LacI proteins are not bound to IPTG at low IPTG levels, leading to an almost complete repression of the P_{lac} promoter due to the high binding affinity between LacI and P_{lac}. Consequently, very few CI proteins are synthesized, and the production of EYFP starting from the λP_{RO12} promoter should be influenced very little by CI repression (Eqs. F7–F9 in Fig. 2) and the steps before it. When the IPTG level rises, the production of EYFP is affected to a higher extent by the CI repression of λP_{RO12}, hence these regulation steps also play important roles. These findings explain the different effects that IPTG variations have on k_{14} to k_{17} sensitivities compared with other rate constants, because k_{14} to k_{17} are the only rate constants that control the transcription and translation of EYFP, whereas all other rate constants are involved in the preceding regulation steps. Since EYFP synthesis is affected by the regulation steps only at high IPTG levels, the sensitivities to the corresponding rate constants such as k_6 and k_{10} to k_{13} increases at elevated IPTG levels. In contrast, the transcription and translation of EYFP stay effective despite IPTG variations, hence the sensitivities of k_{14} to k_{17} are fairly constant.

The laboratory results in Table 1, however, are not in complete agreement with the theoretical predictions in Fig. 4. The RS-HDMR analysis suggests that the steady-state [EYFP] at [IPTG] = 0 should be highly insensitive to k_6, k_{10}, k_{11}, k_{12}, and k_{13}, whereas the fluorescence variations at [IPTG] = 0 in Table 1 are not insignificant among the 16 inverters. This inconsistency can be explained by three possible reasons. First, all the 17 rate constants are picked randomly from large dynamic ranges, and in most of the selections P_{lac} is completely repressed at [IPTG] = 0, leading to minimal CI concentrations and very low [EYFP] sensitivities to mutations on RBS and O_R1. However, we do observe a small population of rate constant sets that results in a moderate expression of CI at [IPTG] = 0 (hence higher sensitivity values to k_6, k_{10}, k_{11}, k_{12}, and k_{13}), which may represent the real system better than most of the random rate constant sets. Since we do not know if those rate constants are more accurate and RS-HDMR generates only statistically averaged results, that small population is overwhelmed by the majority of the rate constant sets. This phenomena immediately raises the need for biosystem identification (discussed at the end of this article), which can provide a much more solid ground for better sensitivity analysis if the rate constant values can be determined with higher precision. Similar to any modeling studies, the inconsistency can also be the result of unmodeled dynamics (e.g., the inherent leakiness of CI or stochastic effects), which leads to the question of how accurate the model needs to be in order to be useful. The model used in the work proved to be reliable in this sense, but one should always be

aware of the danger of overestimating the outcome from a specific model. At last, the possible experimental errors may also contribute to the inconsistency.

The sensitivities of the inverter gain and the slope to the rate constants were also calculated using Eqs. 3 and 5 with X_n replaced by the numerically calculated gain or slope values. The gain g for any rate constant set $\mathbf{k^s}$ is defined as

$$g = [EYFP]_{low}^{*,s}/[EYFP]_{high}^{*,s}, \qquad (6)$$

where $[EYFP]_{low}^{*,s}$ and $[EYFP]_{high}^{*,s}$ represent the steady-state EYFP levels at [IPTG] = 1 μM and 1 mM, respectively. In the simulations, the slope p is determined by

$$p = [EYFP]_{mid+ss}^{*,s}/[EYFP]_{mid-ss}^{*,s}, \qquad (7)$$

where $[EYFP]_{mid}^{*,s} = ([EYFP]_{high}^{*,s} - [EYFP]_{low}^{*,s})/2$ represents the midpoint of the transfer curve and ss is a small step size of [IPTG] ($ss = 1$ μM in this case). A ratio is used to represent the slope because the transfer curves are plotted in logarithmic scale. Due to its sensitivity to laboratory data noise, the slope p of the transfer curves in Fig. 3 is determined by

$$p = [EYFP]_{([IPTG]=100\ \mu M)}/[EYFP]_{([IPTG]=1\ mM)}. \qquad (8)$$

Table 2 shows that both g and p are more sensitive to k_6 than to k_{10}, k_{11}, k_{12}, and k_{13}. Since the true rate constant values are not available, the laboratory data in Fig. 3 could not be used to cross-compare the influence of k_6 on g or p against the other four rate constants. However, the results in Table 2 do indicate that g and p are affected in a similar fashion by these five rate constants, and in Fig. 3, when g is reduced by lowering the k_{10}, k_{11}, k_{12}, and k_{13} values ($a \rightarrow c$ or $b \rightarrow d$), p also decreases. Conversely, when g is affected little by k_6 changes ($a \rightarrow b$ or $c \rightarrow d$), changes in p are also insignificant.

The analysis results in Fig. 4 and Table 2 also reveal that, although k_{14} to k_{17} contribute significantly to [EYFP] variations at all IPTG levels, they are not as important in affecting g and especially p, probably because variations in k_{14} to k_{17} simply shift the whole transfer curve upward or downward, whereas g and p are influenced mostly by the relative [EYFP] difference between high and low IPTG levels. This observation suggests that different optimal mutation targets need to be selected for different circuit optimization purposes. Since the choice of the right targets is not intuitively evident even for this simple circuit, it is very risky (if not impossible) to design experiments on more complex

networks without making the appropriate statistically based sensitivity analyses before performing the mutations.

Caution must be used when making quantitative comparisons between experimental and theoretical results. RS-HDMR is a statistics-based algorithm, which provides sensitivity values averaged over a large dynamical and high-dimensional input space instead of precise sensitivity values for certain rate constants. Because all the experiments start from cells with specific parameter settings, it is anticipated that the experimental and theoretical results will show some discrepancy. However, since the precise values for the model parameters are rarely known in real applications, the RS-HDMR algorithm provides well-defined statistically based information to guide the experiments.

It needs to be emphasized that not all bionetworks are amenable to direct RS-HDMR analysis. The genetic inverter studied in this work belongs to the family of combinational circuits that do not contain any memory elements that may arise from feedback connectivities. In this case, RS-HDMR can be directly executed because the circuit's steady states are uniquely determined by the rate constant values and are independent of the initial concentrations of its components (Katz, 2004; Wakerly, 2003). In contrast, certain biosystems can exhibit multiple steady states, periodic behavior without steady states, or chaotic behavior; and complexity will arise in understanding and/or controlling them. To handle these complex systems, one may divide the circuit responses into qualitatively different subregions and analyze them separately, or devise other appropriate measures of the circuit behavior to avoid the multiplicity in property representations. These possible treatments will be tested in future research.

CONCLUSION

In this article, we introduce a statistics-based, global sensitivity analysis algorithm (RS-HDMR) to better engineer artificial genetic circuits. RS-HDMR estimates the sensitivities of the circuit properties with respect to the circuit model parameters, without knowledge of the precise parameter values. The sensitivity results can then guide the selection of circuit components (e.g., genes and their regulatory components) whose mutations can most effectively optimize specified circuit behavior, thus significantly enhancing the efficiencies of the experiments. As a proof of principle, we measured the properties of 16 pairwise mutants of a genetic inverter and compared them against the RS-HDMR predictions, which confirmed the reliability of the algorithm. The RS-HDMR analysis results indicate that 1), the steady-state EYFP levels are affected differently by mutations in the RBS upstream of the cI coding region and mutations at the O_R1 operator binding site of the λP_{RO12} promoter; 2), the above sensitivities change significantly at different IPTG levels, whereas the sensitivities to the rate constants involved in EYFP transcription and translation show little variation; and 3), the EYFP levels, the gain, and the slope of the

TABLE 2 The first-order (normalized) sensitivities of the gain g and the slope p with respect to several rate constants

	k_6	k_{10}	k_{11}	k_{12}	k_{13}	k_{14}	k_{15}	k_{16}	k_{17}
g	0.94	0.38	0.34	0.10	0.11	0.25	0.26	0.28	0.26
p	0.45	0.21	0.21	0.02	0.02	0.03	0.03	0.03	0.03

transfer curves are affected in different ways by those mutations, which is not intuitively evident and further indicates the need for the RS-HDMR analysis in optimizing different properties of different genetic circuits.

In addition to artificial biocircuit engineering, RS-HDMR can also be employed to construct quantitative models for natural bionetworks (i.e., model parameter estimation/ identification) in a reliable and cost-effective fashion. In most bionetwork identifications, the target system is perturbed first by external elements, such as adding chemicals that upregulate or downregulate certain network components. The response of the network is then recorded and utilized in the model construction. To identify complex bionetworks, it is very important that the most information be obtained from the least number of experiments, as the identification can be very expensive and time-consuming. In a recent work (Feng and Rabitz, 2004), RS-HDMR was successfully applied to estimate the optimal molecular species for perturbing and monitoring a simulated biochemical reaction network, which ensured the effectiveness of the proceeding optimal, closed-loop identification of the network model parameters. Other issues such as the nonlinearity of the bionetwork, data noise, limited number of measurements, and laboratory constraints were also addressed.

Biology is being rapidly transformed into a quantitative data-driven science, fueled by the increasingly powerful technologies in genomics, proteomics, metabonomics, etc. At the same time, biologists have realized that these tools alone cannot easily enhance our understanding of biosystems without the correct incorporation of mathematical tools. This article (and a previous work, Feng and Rabitz, 2004) demonstrates the importance of applying appropriate optimal control techniques in understanding and manipulating complex bionetworks. We are beginning to apply these tools to several systems-biology areas, such as closed-loop learning control of bionetworks (Ku et al., 2004), analysis of neural systems, optimal drug target discovery, and bionetwork connectivity identification, and more importantly, we expect the concept of doing experiments optimally can also be of value in other biological contexts.

REFERENCES

Atkinson, M. R., M. A. Savageau, J. T. Myers, and A. J. Ninfa. 2003. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in *Escherichia coli*. *Cell*. 113:597–607.

Basu, S., D. Karig, and R. Weiss. 2002. Engineering Signal Processing in Cells: Towards Molecular Concentration Band Detection. *Eighth International Meeting on DNA-Based Computers*.

Becskei, A., and L. Serrano. 2000. Engineering stability in gene networks by autoregulation. *Nature*. 405:590–593.

Bower, J. 2001. Computational Modeling of Genetic and Biochemical Networks. MIT Press, Cambridge, MA.

Elowitz, M. B., and S. Leibler. 2000. A synthetic oscillatory network of transcriptional regulators. *Nature*. 403:335–338.

Feng, X., and H. Rabitz. 2004. Optimal identification of biochemical reaction networks. *Biophys. J.* 86:1270–1281.

Francois, P., and V. Hakim. 2004. Design of genetic networks with specified functions by evolution in *silico*. *Proc. Natl. Acad. Sci. USA*. 101:580–585.

Gardner, T. S., C. R. Cantor, and J. J. Collins. 2000. Construction of a genetic toggle switch in Escherichia coli. *Nature*. 403:339–342.

Guet, C., M. Elowitz, W. Hsing, and S. Leibler. 2002. Combinatorial synthesis of genetic networks. *Science*. 296:1466–1470.

Hasty, J., J. Pradines, M. Dolnik, and J. J. Collins. 2000. Noise-based switches and amplifiers for gene expression. *Proc. Natl. Acad. Sci. USA*. 97:2075–2080.

Hasty, J., D. McMillen, F. Isaacs, and J. J. Collins. 2001a. Computational studies of gene regulatory networks: *in numero* molecular biology. *Nat. Rev. Genet.* 2:268–279.

Hasty, J., F. Isaacs, M. Dolnik, D. McMillen, and J. J. Collins. 2001b. Designer gene networks: towards fundamental cellular control. *Chaos*. 11:207–220.

Hasty, J., D. McMillen, F. Isaacs, and J. J. Collins. 2001c. Computational studies of gene regulatory networks: *in numero* molecular biology. *Nat. Rev. Genet.* 2:268–279.

Hasty, J., D. McMillen, and J. J. Collins. 2002a. Engineered gene circuits. *Nature*. 420:224–230.

Hasty, J., M. Dolnik, V. Rottschafer, and J. J. Collins. 2002b. A synthetic network for entraining and amplifying cellular oscillations. *Phys. Rev. Lett.* 88:148101-1–148101-4.

Jong, H. D. 2002. Modeling and simulation of genetic regulatory systems: a literature review. *J. Comput. Biol.* 9:67–103.

Katz, R. H. 2004. Contemporary Logic Design. Addison-Wesley, Boston, MA.

Ku, J., X. Feng, and H. Rabitz. 2004. Closed-loop learning control of bionetworks. *J. Comput. Biol.* In press.

Lewin, B. 2000. Genes VII. Oxford University Press, New York.

Li, G., C. Rosenthal, and H. Rabitz. 2001. High dimensional model representations. *J. Phys. Chem. A*. 105:7765–7777.

Li, G., S.-W. Wang, H. Rabitz, S. Wang, and P. Jaffe. 2002. Global uncertainty assessments by high dimensional model representations (HDMR). *Chem. Eng. Sci.* 57:4445–4460.

McAdams, H. H., and A. Arkin. 1998. Simulation of prokaryotic genetic circuits. *Annu. Rev. Biophys. Biomol. Struct.* 27:199–224.

Miller, J. H., and W. S. Reznikoff. 1980. The Operon. Cold Spring Harbor Laboratory, New York.

Saltelli, A., K. Chan, and E. M. Scott. 2000. Sensitivity Analysis. John Wiley & Sons, New York.

Wakerly, J. F. 2003. Digital Design: Principles and Practices. Prentice Hall, Englewood Cliffs, NJ.

Wang, S.-W., P. G. Georgopoulos, G. Li, and H. Rabitz. 2003. Random sampling-high dimensional model representation (RS-HDMR) with nonuniform distributed variables: application to an integrated multimedia/multipathway exposure and dose model for trichloroethylene. *J. Phys. Chem. A*. 107:4707–4716.

Weiss, R. 2001. Cellular computation and communications using engineered genetic regulatory networks. PhD thesis, MIT, Cambridge, MA.

Weiss, R., and S. Basu. 2002. The Device Physics of Cellular Logic Gates. *First Workshop on Non-Silicon Computing*. Boston, MA.

Yildirim, N., and M. C. Mackey. 2003. Feedback regulation in the lactose operon: a mathematical modeling study and comparison with experimental data. *Biophys. J.* 84:2841–2851.

Yokobayashi, Y., R. Weiss, and F. H. Arnold. 2002. Directed evolution of a genetic circuit. *Proc. Natl. Acad. Sci. USA*. 99:16587–16591.