

## Review Article

# Mathematical Modeling: Bridging the Gap between Concept and Realization in Synthetic Biology

**Yuting Zheng and Ganesh Sriram**

*Department of Chemical and Biomolecular Engineering, University of Maryland, 1208D,  
Chemical and Nuclear Engineering Building 090, College Park, MD 20742, USA*

Correspondence should be addressed to Ganesh Sriram, [gsriram@umd.edu](mailto:gsriram@umd.edu)

Received 21 December 2009; Accepted 7 March 2010

Academic Editor: Hal Alper

Copyright © 2010 Y. Zheng and G. Sriram. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Mathematical modeling plays an important and often indispensable role in synthetic biology because it serves as a crucial link between the concept and realization of a biological circuit. We review mathematical modeling concepts and methodologies as relevant to synthetic biology, including assumptions that underlie a model, types of modeling frameworks (deterministic and stochastic), and the importance of parameter estimation and optimization in modeling. Additionally we expound mathematical techniques used to analyze a model such as sensitivity analysis and bifurcation analysis, which enable the identification of the conditions that cause a synthetic circuit to behave in a desired manner. We also discuss the role of modeling in phenotype analysis such as metabolic and transcription network analysis and point out some available modeling standards and software. Following this, we present three case studies—a metabolic oscillator, a synthetic counter, and a bottom-up gene regulatory network—which have incorporated mathematical modeling as a central component of synthetic circuit design.

## 1. Introduction

Synthetic biology aims to design novel biological circuits for desired applications, implemented through the assembly of biological parts including natural components of cells and artificial molecules that emulate biological behavior [1, 2]. Because of its parts-to-whole approach, synthetic biology has a significant engineering component. Engineering endeavors typically involve the three classical engineering strategies: standardization (ensuring that components of a system are compatible and exchangeable), decoupling (dissecting a system into less complicated subsystems), and abstraction (streamlining a problem to focus only on the pertinent facets) [3–5]. It may appear that it should be possible to apply these strategies toward constructing a synthetic biological circuit in a manner similar to constructing an electric or electronic circuit. The attainment of this ideal goal is, however, impeded by the overwhelming complexity of biological systems with their myriad biomolecules and interconnections as well as sparse databases of gene function [3]. Consequently it is challenging to convert design concepts to predicted results.

This stumbling block in synthetic biology can be alleviated by the use of computer-aided mathematical modeling. Modeling is a powerful and often indispensable link between design and realization in engineering. It can predict the dynamics of a network under several different conditions and combinations thereof. Due to this, a user can search large parameter spaces *in silico* to identify the small regions of parameter space that produce the desired behavior or the most effective design or, alternatively, avoid parameter values that result in undesired responses. Modeling also provides the capability of using knowledge about the constituent parts of a system to predict the behavior of a system as a whole. Therefore, mathematical modeling serves as a bridge connecting a conceptual design idea to its biological realization (Figure 1).

In this review we present mathematical modeling concepts as relevant to synthetic biology and illustrate their application through the discussion of three case studies [6–8]. While the role of modeling in synthetic biology has been expertly reviewed before (e.g., [4, 9, 10]), this review aims to build upon the previous reviews by collecting

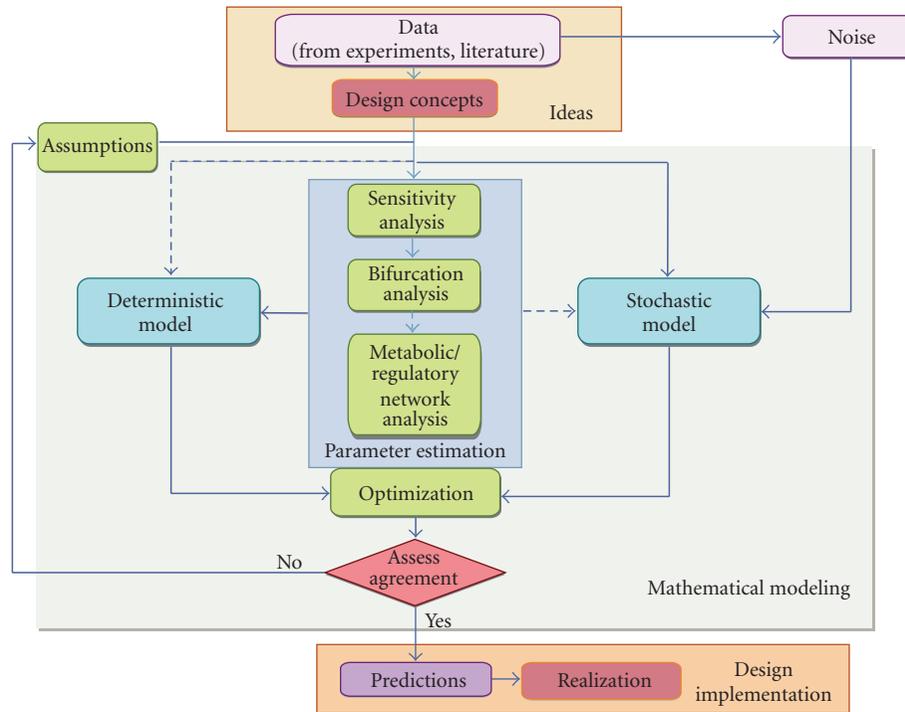


FIGURE 1: The role of mathematical modeling in synthetic biology. Computer-aided mathematical modeling bridges a design concept to realization in synthetic biology. Solid lines depict typical steps that have to be performed while developing a model; dashed lines depict unusual scenarios or conditions under which the steps shown by the corresponding solid lines are trivial or can be bypassed. A concept or ideas for designing a circuit for a particular function may be inspired by data from experiments or the literature. A mathematical model is then formulated on the basis of certain assumptions. The framework of a model could be deterministic or stochastic. The development of a model generally begins with the estimation of parameters that govern the model; this is a process that involves sensitivity analysis, bifurcation analysis, and, under certain circumstances, metabolic and transcription (regulatory) network analysis. The dashed line from design concepts to deterministic model indicates that, in some cases, parameter estimation is trivial or can be bypassed for this type of model. A stochastic model is developed by employing statistical functions to mimic system dynamics and considering fluctuations in the data. The dashed line from parameter estimation to stochastic model indicates that in some cases, parameter estimation may offer information in choosing statistical functions when constructing a stochastic model. Optimization is required for both models and is complete when the model exhibits an agreement (goodness of fit) with experimental data. A good agreement enables reliable prediction of system behavior and further biological realization, whereas unsatisfactory agreement requires the revision of the initial assumptions and the beginning of the next modeling cycle. See text and Figure 2 for explanations of terms.

several modeling methodologies into a single article and exemplifying them with in-depth case studies.

Figure 1 summarizes the important role played by mathematical modeling in synthetic biology and the key steps in the modeling process. Briefly, a model is formulated on the basis of certain assumptions about the system. Two broad types of modeling frameworks are available: deterministic modeling and stochastic modeling. Depending on the model type and the system, the estimation of parameters in the model could be a crucial step in obtaining a satisfactory model, and optimization may play a major role in this process. The predictions of a completed model are used, in conjunction with sensitivity analysis, bifurcation analysis, or, in certain cases, metabolic and regulatory network analyses to obtain insights toward an effective design. Below, we present a detailed discussion of the steps involved in modeling.

## 2. Assumptions Underlying a Model

Biological systems are difficult to model and simulate despite a wealth of data on the structure and function of biomolecules and on cellular mechanisms. This is because biological systems exhibit complexity on several scales. Firstly metabolites, metabolic fluxes, proteins, RNA, and genes network in a highly complex manner; furthermore, their interconnections could constitute feedback or feed-forward loops that respond at various time scales [11]. Secondly, living systems can be highly sensitive to time-variant environmental conditions such as light, humidity, and supply of nutrients. These and other unknown causes of uncertainty result in “biological errors”, which are distinct and usually greater in magnitude than instrumentation or measurement errors. Therefore, it is difficult to exactly predict the output of a biological system as compared to

### Mathematical modeling definitions and techniques

**Agreement.** See goodness of fit.

**Assumption.** A presupposition that forms the basis of a model and trims down the complexity of the model. Examples are spatial or cell population homogeneity, lack of intracellular compartmentation or the existence of equilibrium, steady state or quasi-steady state.

**Bifurcation analysis.** An analysis that identifies the boundaries between the regions in parameter space that result in drastically different system dynamics, e.g. stable versus oscillatory. Two popular methods of bifurcation analysis are saddle node bifurcation and Hopf bifurcation.

**Deterministic model.** A model that endeavors to mimic a real system with analytical equations (usually ODEs or PDEs) that include numerical parameters. The output of such a model is predictable and reproducible (see Example, right).

**Goodness of fit.** A criterion used to judge the conformity between experimental data points and simulated points, using which parameter values can be accurately estimated. A chi-squared statistic is commonly used as a goodness of fit criterion. Agreement assessment involves statistical methods such as  $t$  tests and confidence interval estimation.

**Metabolic flux analysis.** Quantification of carbon flow through a network of metabolic pathways using stoichiometry, carbon atom rearrangements and isotopomer data from nuclear magnetic resonance and mass spectrometry.

**Noise.** Unpredictable fluctuation in the variables that describe the state of the system. Deterministic models usually do not consider noise, where as it is a significant component of stochastic models.

**Optimization.** A method to locate the minimum or maximum of an objective function. Local optimization searches for the optimum within a local neighborhood where as global optimization searches for the best optimum in the constrained parameter space.

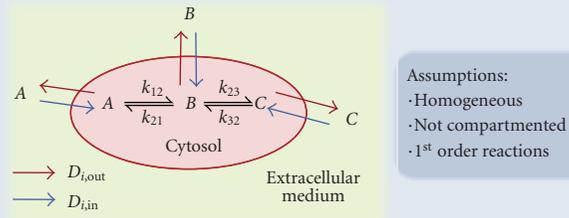
**Sensitivity analysis.** The study of how the variations of certain parameters in a constrained domain will affect the response of a model; it enables the identification of parameters that are crucial to the model.

**Stochastic model.** A model that endeavors to mimic a real system with equations and parameters that vary stochastically. Such a model incorporates fluctuations inherent in real systems.

**Transcription network analysis.** An analysis of the relationships between genes and their regulators (usually transcription factors). Such an analysis aims to identify regulatory motifs underlying observed phenotypes.

### Example of a deterministic model

Consider the development of a deterministic model for the following simple network.



Assumptions:  
 · Homogeneous  
 · Not compartmented  
 · 1<sup>st</sup> order reactions

### Model equations

$$\frac{dA}{dt} = 0 \cdot A + k_{12}B + 0 \cdot C + D_1$$

$$\frac{dB}{dt} = k_{21}A + 0 \cdot B + k_{23}C + D_2$$

$$\frac{dC}{dt} = 0 \cdot A + k_{32}B + 0 \cdot C + D_3$$

### Parameters

$i, j = 1, 2, 3$

$k_{ij}$ : reaction rate constants

$D_i$ : the rate of transport of each metabolite in/out of the cell

$$D_i = X_{i,in} - X_{i,out}$$

These equations can be written in matrix form:

$$\begin{bmatrix} \frac{dA}{dt} \\ \frac{dB}{dt} \\ \frac{dC}{dt} \end{bmatrix} = \begin{bmatrix} 0 & k_{12} & 0 \\ k_{21} & 0 & k_{23} \\ 0 & k_{32} & 0 \end{bmatrix} \begin{bmatrix} A \\ B \\ C \end{bmatrix} + \begin{bmatrix} D_1 \\ D_2 \\ D_3 \end{bmatrix} \Rightarrow \frac{dX}{dt} = F(N, t; \theta)$$

Where  $X$  and  $N$  are vectors comprised of concentrations of the species  $A, B$  and  $C$  (and are identical in this case),  $dX/dt$  is the rate of change of  $X$ ,  $\theta$  is a vector of model parameters.

With  $D_i$  measurable, this model can be solved for the time course trajectories of the metabolites  $A, B$  and  $C$ . The steady state(s) of the model can be analyzed by setting  $dX/dt = 0$ .

FIGURE 2: Terms used in mathematical modeling and an example of a simple (deterministic) mathematical model.

a mechanical or an electrical system and one has to often reconcile with an approximate reproduction.

However, a biological system can often be simplified to a level that permits a user to obtain insights toward synthetic circuit construction [11]. For example, Ma'ayan et al. [12] demonstrated how simplifying the dynamics of single components could lead to valuable information on a system's function. Simplification of a model requires the making of various assumptions. A commonly used assumption is homogeneity, both within the cell and within a cell population. Spatially homogeneous time-variant systems can be modeled by ordinary differential equations (ODEs) (equation (1), see Figure 2, e.g.). However, time-variant systems that feature compartmentation [13], spatial segregation, or intracellular gradients [14] may require the use of

partial differential equations (PDEs). Although solving PDEs (and thereby non-homogenous models) is computationally much more intensive than solving ODEs, it can pay off well. For example, effects such as the spatial segregation of two enzymes that may generate intracellular gradients or the effect of protein diffusivity on enzyme activity can be simulated by using nonhomogenous models [14]. Closely related to spatial homogeneity is the assumption of cell population homogeneity, which is very frequently employed in models of biological systems. However, the modeling of heterogeneous populations in chemical reactors [15] has found application in the modeling of heterogeneous cell populations, and stochastic models frequently employ it. Besides the homogeneity assumption, most models involving enzyme kinetics or transcriptional regulations also assume

equilibrium [8], steady state [7], or quasi-steady state [6]. Such an assumption can remove time-dependence from the model and converts ODEs to simpler algebraic equations. The task of formulating the assumptions underlying a model is a fine balancing act between trimming down the complexity of the system while retaining the features of the system that are crucial to making reliable predictions for the application at hand. If a model based on certain assumptions does not agree with experimentally observed behavior, then the assumptions have to be revised [9]. This makes mathematical modeling an iterative process (Figure 1).

### 3. Types of Model Frameworks

**3.1. Deterministic Mathematical Models.** Mathematical models of biological systems can be categorized into two major types: deterministic and stochastic [4]. A deterministic model emulates a real system with analytical equations (usually ODEs or PDEs) that include numerical parameters. These equations are usually mass balances on cellular species (Figure 2), and the state of the system that is predicted by such a model is reproducible [4]. In contrast, a stochastic model endeavors to represent a real system with randomly interacting particles or species. The rate of each reaction between the species follows a probabilistic equation [9]; additionally, the time between the reactions can also vary. Stochastic models usually incorporate the fluctuations and noise inherent in real biological systems and examine the effect of noise on system dynamics [4].

Figure 2 explains the construction of a deterministic model for a simple network. Deterministic models usually employ differential equations used to describe interactions or reactions between biomolecules:

$$\frac{dX}{dt} = F(N, t; \theta), \quad (1)$$

where  $X$  and  $N$  are vectors comprised of species concentrations (and may be identical),  $dX/dt$  is the rate of change of  $X$ ,  $\theta$  is a vector of model parameters (see the following section on model parameters), and  $F(N, t; \theta)$  is a nonlinear vector function that relates rates of change to concentrations [4]. Dynamic simulations of a system modeled by equation (1) are quite straightforward and will reveal the time-dependent characteristics of the system by generating time series trajectories of the species concentrations. Furthermore, the simulations help to analyze the behavior of a network when feedforward or feedback regulations are integrated into it. Such analysis has shown that the dynamic properties of feedforward loops depend on their specific architecture [16, 17], that positive or double negative feedbacks often introduce ultrasensitive or bistable switches [18, 19], and that negative feedbacks may reduce instability of the system [20]. Furthermore, multiple steady states or oscillations may occur as the result of positive and negative feedbacks [21, 22].

To analyze steady states of a time-dependent biological system, the time derivatives in (1) are set to zero [4]:

$$F(N, t; \theta) = 0, \quad (2)$$

which represents the steady state(s) of the system, is usually combined with bifurcation analysis to obtain the range of parameters in which the system will exhibit certain desired behaviors such as oscillations [23] (see Section 5 below).

Numerous deterministic models have been developed for biological systems, including several for synthetic circuits. Case studies I and II discussed in this article [6, 7] employ deterministic models.

**3.2. Stochastic Mathematical Models.** In deterministic models, every interaction and every parameter value is certain. Therefore, such models predict identical system dynamics for the same set of parameter values and initial conditions. However, real systems are characterized by unexpected and irreproducible fluctuations. To capture these fluctuations and their consequences on the behavior of the system, an alternative type of model, the stochastic model, is used. Such models mimic a system as a collection of interacting particles, with the reaction rates being governed by probabilistic rate laws [9]. An example of such a rate law is the chemical Master equation [10]. Stochastic simulation algorithms (SSAs) [10] such as Gillespie's algorithm are then used to simulate the state of the system.

One approach in stochastic modeling is to assume that a system is comprised of randomly interacting biomolecules, wherein the reactions between the molecules are modeled as Poisson processes with a probabilistically determined rate parameter [9]. Another approach is to perceive a time-variant system as a discrete time stochastic process. This approach uses a random variable or a vector  $X_n$  to indicate the discrete state of the system amongst several (finite or infinite) possible states [24]. The fewer the system states, the easier it is to construct a stochastic model. Guido et al. [8] have employed the latter approach with six system states to develop a stochastic model for a bottom-up gene regulatory network. In this type of stochastic model, the probability  $p_i$  at time  $n$  of each system state  $S_i$  is computed based on certain assumptions [24]:

$$p_i(n) = P(X_n = i). \quad (3)$$

System responses or outputs such as the rate of synthesis of green fluorescent protein (GFP) are then described in terms of the state probabilities:

$$y = \sum_{i=1}^n g_i p_i, \quad (4)$$

where  $y$  is the net output resulting from a combination of  $n$  states and  $g_i$  is the synthesis rate contributed by each state  $S_i$ . The probability of a system state  $p_i(n)$  could be estimated by taking into account noises from synthesis and degradation of mRNA, GFP, or transcription factors and physical properties of the cell or system [8]. Finally, equation-free stochastic models have also been developed [25].

Several stochastic models have been developed for synthetic biological circuits and related simple biological systems [8, 25–29]. In case study III of this article, we discuss a stochastic model for a bottom-up gene regulatory network reported in Guido et al. [8].

#### 4. Parameters in a Model

Any model contains several variables that do not represent the system state, but whose values govern the dynamics of the equations in the model. Such variables include reaction rate constants, equilibrium constants, diffusivities, and other physical properties. These are termed “parameters” of the model, as opposed to “state variables” such as species concentrations that represent the state of the system. To make useful predictions from a model, the parameters in the model have to be accurately estimated.

Mechanistic models, which are based on physical and chemical laws, include parameters that carry physical, chemical, or biological meaning. However, there could be many instances where not much information is available about a system, and constructing a “black box” model is the only option available. The parameters of such a model do not carry physical or biological meaning, but their estimation is nevertheless indispensable to the success of the model. Occasionally, information about a system could be so meager that even a black box model cannot be constructed. In such cases, a reverse engineering approach is employed to translate observable information to not only parameters but also model equations. This approach involves searching through (discrete) topological space instead of (continuous) numerical parameter space. Sometimes, combining the topological and numerical parameters for a system and simultaneously searching for both types of parameters has many advantages in understanding systems that are sparsely known [30]. Sometimes, parameter estimation can be largely bypassed. Recently, Tran et al. [31] described an approach called ensemble modeling, suited to modeling metabolic networks. This technique bypasses the requirement of obtaining accurate parameter values by reducing an initial set of models. The final set of models so obtained is capable of describing phenotypes of enzyme perturbations. Such approaches offer versatile ways to attack parameter problems in modeling.

**4.1. Parameter Estimation.** Parameter estimation is known as the “inverse problem” or “model calibration” and is both a key step and a limiting step of model construction [32, 33]. Parameter estimation typically leads to a first working model. If this initial model exhibits significant departures from experimental data (a frequent occurrence), further experiments may need to be performed to refine the parameter values. This process is repeated iteratively until a satisfactory model is constructed [34, 35]. Model parameters can occasionally be found from the literature or estimated manually, although this is usually feasible for small parameter spaces and simple systems. In general, parameter estimation from experimental observations requires sophisticated techniques such as described below. Amongst these, (global) optimization, a computationally intensive but powerful and robust method, is widely used.

**4.2. Optimization.** Parameter estimation is generally an optimization problem that involves locating the optimum (minimum or maximum) of an objective function that

represents how well the model simulations agree with experimental data. This can be expressed as

$$\min_{\theta} \Phi(\theta). \quad (5)$$

The function  $\Phi(\theta)$ , which represents the goodness of fit between experiment and simulation, is a scalar function of the parameter vector  $\theta$ . Its optimum is determined by iteratively adjusting the values of the components of  $\theta$  and sometimes revising model assumptions [4]. The function  $\Phi(\theta)$  is most frequently a weighted sum-of-squares error (a chi-squared statistic) between the experimental data points and the corresponding simulated points [36]. Other objective functions such as the Bayesian estimator and the maximum likelihood estimator also work well [33]. The process of adjusting and refining parameter values to reach the optimum of this objective function can be performed manually for linear or piecewise linear models [37]. However, many biological processes are not only nonlinear but also random, thus necessitating nonlinear models, stochastic models, or both [37–39]. In such cases, several general methods of parameter estimation are available, including some that are particularly suited to nonlinear dynamic systems typical in biology [40–43]. Even while using sophisticated algorithms, parameter estimation can involve unexpected complications such as the inability of a given optimization algorithm to effectively search a parameter space. In such cases, an exhaustive searching of parameter space can sometimes be accomplished well by stochastic Monte Carlo algorithms [44]. However, the computation involved in such exhaustive searching may often become prohibitively expensive when the number of parameters runs into hundreds or thousands. Additionally, unobserved or unknown interactions that were not accounted for in the formulation of the model can result in unsuccessful parameter searches and will require the model assumptions to be revisited.

The calculation of the first and second partial derivatives of the objective function is sometimes useful in optimization. Gradient search optimization algorithms depend on the partial derivatives of the objective function (or the partial derivative matrix, the Jacobian) for their success [4]. The Jacobian  $J$  is defined as

$$J = \frac{\partial \Phi}{\partial \theta}. \quad (6)$$

Occasionally it is also useful to analyze how the combination of two parameters may affect the system dynamics; this is accomplished through the Hessian matrix  $H$ :

$$H = \frac{\partial^2 \Phi}{\partial \theta \partial \theta^T}, \quad (7)$$

which is a matrix containing the second partial derivatives of the objective function with respect to pairs of parameters.

In gradient search methods it is not always possible to reach the global optimum of the objective function, especially for nonlinear objective functions that may have several local optima far away from the global optimum.

Therefore, it may become necessary to sacrifice the speed of the gradient search methods for the exhaustive searching abilities of probabilistic methods [45]. Such methods avoid the inferior solutions often found by gradient search methods [46]. Examples of probabilistic optimization methods are simulated annealing, genetic, and evolutionary algorithms [47, 48] or scatter searches [49]. Conversely, a local minimum may sometimes be sufficient to generate parameter values that are practical. In some particularly difficult combinatorial optimization problems the local minimum found near the global minimum may turn out to be a better choice [50]. On some rare occasions, perhaps many of the optima could be useful for the researcher.

## 5. Model Analysis Techniques

After a satisfactory model is constructed, the analysis of the model and its predictions provides crucial input toward designing synthetic circuits that exhibit a given behavior. Here, we discuss two analysis techniques: sensitivity analysis and bifurcation analysis.

**5.1. Sensitivity Analysis.** Sensitivity analysis, which analyzes how sensitive a system is with respect to changes in parameter values [51], is useful in quantifying the significance of a parameter or parameters on system performance [4]. Local sensitivity analysis analyzes the effects of small perturbations whereas global sensitivity analysis is used to analyze the effects of perturbing parameter values over the entire parameter ranges. Caution should be exercised while extrapolating the results of sensitivity analyses to an operating point far away in parameter space, as this may not accurately predict system behavior at the operating point.

For local sensitivity analysis a sensitivity  $s$  is defined as

$$s = \frac{\partial G(N)}{\partial \theta}, \quad (8)$$

where  $N$  is a vector comprised of species concentrations,  $G(N)$  represents a system state or output, and  $\theta$  is a vector of parameters. The magnitudes of the elements of the vector  $s$  are proportional to the effect of the corresponding parameters [4]. In global sensitivity analysis, the parameter space (constrained by physical limitations, mass fraction summations, etc.) is explored by methodologies such as random sampling-high dimensional model representation [52], multiparametric sensitivity analysis, or Monte Carlo simulation [53].

**5.2. Bifurcation Analysis.** Bifurcation analysis is crucial to understanding and analyzing steady states, oscillations, and other dynamic features of a system and has found use in numerous modeling studies [6, 54, 55]. In many nonlinear models, the parameter space can be divided into regions that lead to a stable system, an unstable system, or a periodic (oscillatory) system. Identifying the boundaries of these regions will enable the design of a synthetic circuit that exhibits desired behavior. Two popular methods of bifurcation analysis are saddle node bifurcation analysis and

Hopf bifurcation analysis. Saddle node analysis endeavors to investigate the threshold where a system functions as a biological switch and thus separates the region of the parameter space that confers monostability [56]. Hopf bifurcation analysis is predominantly used to analyze oscillators and can characterize the critical parameter values that enable a system to transition from a stable steady-state solution to a periodic solution [57, 58]. In Hopf bifurcation analysis the eigenvalues of the Jacobian matrix, equation (6), are used to obtain the threshold parameter values at which the system's behavior changes drastically. A phase diagram is then constructed by identifying the points where the real parts of a pair of complex conjugate eigenvalues crosses zero while all other eigenvalues have negative real parts (see [6], e.g.; this paper is discussed in case study I below). While there are numerous examples of bifurcation analysis of deterministic models, one work [27] has comprehensively treated the applications of bifurcation analysis to a stochastic model of an oscillatory system.

## 6. Modeling as a Part of Phenotype Analysis

Synthetic biology can also benefit from metabolic flux and transcription network analyses, which combine high-throughput experimental observations (such as metabolome, isotopomer, and gene expression profiles) with mathematical modeling to quantitatively describe the phenotype of a biological system. This type of analysis could be particularly useful for highly complex systems. For example, Noirel et al. [59] presented a probabilistic metabolic model that was useful in analyzing the systemic metabolic effects of inserting synthetic circuits into a cell. Of specific relevance to synthetic biology is an elegant work [60] that used optimization to identify how regulation could be superposed on a metabolic network to optimize the network.

**6.1. Metabolic Pathway Analysis.** Isotope-assisted [61–63] metabolic flux analysis is a powerful tool to evaluate carbon flow in metabolic networks and could be relevant in synthetic biology. In this method, material balance models for cellular species are used together with measurements of extracellular metabolites or isotopomers (from nuclear magnetic resonance or mass spectrometry) to obtain metabolic flux maps of a system. While the mathematical modeling approaches discussed above are generally still valid in flux analysis, these models can be enormously complex due to the vast numbers of reactions, myriad carbon atom rearrangements, reaction reversibilities, and variations of intracellular fluxes in real time. This complexity is significantly reduced through the use of compartmental matrix techniques such as cumulative isotopomers (cumomers) [64], bond isomers (bondomers) [65, 66], or elementary metabolite units [63]. Despite this reduction in complexity, metabolic flux analysis of compartmented [61] or instationary systems [62, 67] requires tremendous amounts of computation.

Another set of powerful techniques for modeling metabolic networks includes flux balance analysis (FBA) [68–70] or genome-scale metabolic modeling [71–73] and

TABLE 1: Modeling techniques used in the case studies.

Case study	Model types		Modeling techniques			
	Deterministic	Stochastic	Parameter estimation	Optimization	Sensitivity analysis	Bifurcation analysis
Case study I Fung et al. [6]; designed a synthetic oscillator in <i>E. coli</i> .	×		×	×	×	×
Case study II Friedland et al. [7]; developed synthetic counters in <i>E. coli</i> that can count up to two or three induction events.	×		×	×	×	
Case study III Guido et al. [8]; constructed regulatory networks by assembling simpler building blocks.	×	×	×	×	×	

ensemble modeling [31]. In FBA, a metabolic network is modeled with linear stoichiometric equations, constrained by factors such as extracellular flux measurements and reaction irreversibilities. This model is usually solved by linear optimization and results in a map of steady-state flux values. Through such a reconstruction of a metabolic network, FBA can offer important insights towards selecting gene deletion targets.

Another valuable technique, metabolic control analysis [74–76], aims to elucidate the interdependence of various parts of a metabolic network. The outcomes of this technique are metrics such as flux control coefficients [77], which represent the amount of control exerted by one system component (such as a metabolite) on another system component (such as an enzyme). This method has much to offer toward the important problem of linking genome and phenotype [78].

**6.2. Transcription Network Analysis.** Determining how genes are controlled by regulatory motifs is an important problem in biology. Because synthetic circuits are composed of well-characterized components, they can be used to investigate and quantify transcription networks. Such an investigation would employ a combinatorial technique to construct a circuit comprised of numerous genes and a smaller number of regulatory motifs [79]. The high-dimensional output of such a network is gene expression data and is the end product of the low-dimensional regulatory signals (transcription factor activities) and the strengths of the connectivities between the transcriptional motifs and genes [80]. These transcription factor activities and connectivities are quantified by analyzing the measured gene expression data, using one or more of several available methods [81]. These methods include principal component analysis [82], singular value decomposition [83, 84], independent component analysis [85], network component analysis [80], or state-space models [86]. Network component analysis is a powerful method that uses a priori knowledge about connectivities between transcription factors and genes together with gene expression data to quantitatively infer transcription factor activities and the strengths of the transcription factor-gene

connectivities. The a priori information is obtained from databases or experimental techniques such as ChIP-chip analysis [87].

## 7. Modeling Standards and Software

Several standards and software are now available to simplify the process of building mathematical models and thereby bridge the gap between model description and prediction of the system’s behavior. System biology markup language (SBML) [88] and synthetic biology open language (SBOL) (<http://dspace.mit.edu/handle/1721.1/49523>) are two examples of standards. Both are computer-readable formats for representing models and facilitate the sharing of models between researchers and between different software platforms. Several modeling software are available to practitioners of synthetic biology. These software, which are usually written in popular computer languages such as C++, feature a user interface, relatively simple ways to input information and graphical output of the modeling outcomes, thus relieving users of the burden of setting up and solving mathematical equations. A nonexhaustive listing of these software includes Athena (<http://www.codeplex.com/athena>), BioJade (<http://web.mit.edu/jagoler/www/biojade>), Gepasi (<http://www.gepasi.org>), SynBioSS (<http://synbioss.sourceforge.net/>), which reads and writes in SBML, and TinkerCell (<http://www.tinkercell.com/Home>) which enables users to incorporate new features through custom programs written in C or Python.

## 8. Case Studies

Below we present three case studies that illustrate several of the previously discussed modeling methodologies. The case studies feature two deterministic models [6, 7] and one stochastic model [8]. Table 1 summarizes the principal modeling techniques used in the case studies.

**8.1. Case Study I: Deterministic Model of a Synthetic Oscillator.** Fung et al. [6] reported the mathematical model-aided design of a gene-metabolic oscillatory circuit called

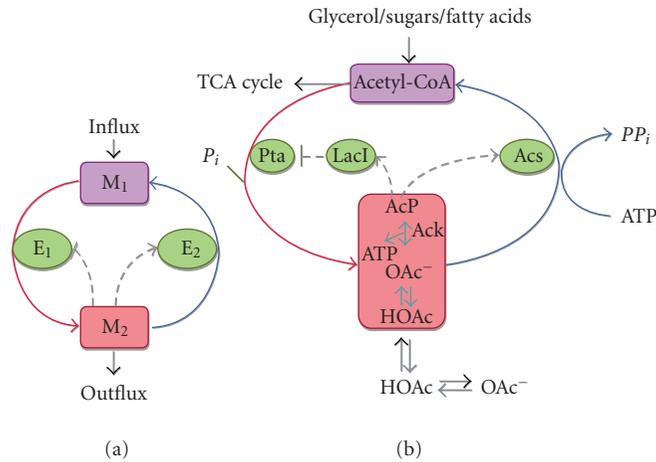


FIGURE 3: Conceptual design and biological implementation of the oscillatory circuit metabolator in Fung et al. [6]. (a) Conceptual design. The metabolator consists of two interconverting metabolite pools  $M_1$  and  $M_2$ ; their interconversions are catalyzed by the enzymes  $E_1$  and  $E_2$ . Dashed lines indicate positive (arrow) and negative (blunt bar) regulation by  $M_2$  at the transcriptional or translational level; the accumulation of  $M_2$  represses  $E_1$  and induces  $E_2$ . The circuit functions as follows. Influx into the circuit (from upstream processes) increases the concentration of  $M_1$ , which is converted to  $M_2$  by  $E_1$ . Initially the concentration of  $M_1$  is high and  $M_2$  is low. However,  $M_2$  gradually accumulates causing  $E_1$  to be repressed and  $E_2$  to be induced, eventually causing a net conversion of  $M_2$  to  $M_1$ , which then starts a new cycle. (b) Biological implementation. The design of the metabolator was implemented using the acetate pathway in *E. coli*. The  $M_1$  pool is acetyl-CoA; the  $M_2$  pool consists of AcP,  $OAc^-$ , and HOAc. Pta and Acs correspond to enzymes  $E_1$  and  $E_2$ . Pta converts Acetyl-CoA to AcP, and AcP is further converted to  $OAc^-$  by Ack. The protonated form of  $OAc^-$  (HOAc) is permeable across the cell membrane. AcP is used as a signaling molecule and functions as follows. When AcP builds up, it will activate promoter *glnAp2* through phosphorylation. The promoter *glnAp2* controls the expression of Acs and lac repressor (LacI), and LacI in turn represses the expression of Pta. Ack: acetate kinase; AcP: acetyl phosphate; Acs: acetyl-CoA synthetase; HOAc: protonated form of acetate; LacI: lac repressor;  $OAc^-$ : acetate; Pta: phosphate acetyltransferase (adapted from Fung et al. [6]).

the metabolator. Toward designing an oscillatory circuit the authors conceived a network with two interconverting metabolite pools wherein one metabolite differently regulates the enzymes that interconvert the two pools. Such a network is theoretically capable of oscillation (see Figure 3(a)). The circuit was implemented in *Escherichia coli* using the acetate pathway (see Figure 3(b)). Under certain circumstances, the sizes of the pools  $M_1$  (Acetyl-CoA) and  $M_2$  (lumped pool of Ack, AcP,  $OAc^-$ , and HOAc) can oscillate. The readout of this network was a GFP, located downstream of the network such that the oscillations of the GFP readout reflect any oscillations occurring in the network. A mathematical model was developed to analyze the behavior of the metabolator and determine the conditions under which oscillations would occur. The model was deterministic and employed ODEs of the following form:

$$\frac{dX}{dt} = V_{in} - V_{out}, \quad (9)$$

$$\frac{dX}{dt} = \text{synthesis rate} - \text{degradation rate}, \quad (10)$$

where  $X$  represents any pool and  $V_{in}$ ,  $V_{out}$  indicate its influx and outflux, respectively. Equation (9) was used to describe the metabolite pools  $M_1$  and  $M_2$  while equation (10) and Michaelis-Menten rate laws were used to describe the kinetics of the enzymes driving the  $M_1$ - $M_2$  interconversions. Using parameter values or ranges typical for this system, the authors implemented the model with a fourth-order Runge-

Kutta algorithm. Parameter sensitivity analysis showed that increasing glycolytic flux increases the oscillatory capability of the system (Figure 4(a)). This prediction was tested by experimentally varying the glycolytic flux through the feeding of different carbon sources (glucose, fructose, mannose, and glycerol), each of which resulted in a different value of this flux. An explanation for this observation is that there existed a threshold value of the glycolytic flux beyond which the system would oscillate. Conversely, high external acetate was predicted to suppress oscillation (Figure 4(c)), which was also verified by experiments.

Hopf bifurcation was then used to characterize the dynamics of the model and determine the transition point at which the steady state would turn to a periodic state. Figure 4(b) depicts the phase diagrams constructed by mapping the locus of Hopf bifurcation. The oscillations approach a limit cycle and were stable, as determined through Floquet analysis. As previously inferred, oscillations occur above a threshold glycolytic flux value and are not sustained at a high acetate concentration. Furthermore, by comparing the modeling simulations with experimental data, the authors found that the inherent noise in gene expression was an important determinant of the amplitude of oscillation.

This work represents a universal approach to construct a synthetic biological circuit with interesting dynamics and beautifully demonstrates the key role played by mathematical modeling in realizing a design concept. Modeling offered valuable insights on the analysis of experiment data and made nontrivial predictions of the system dynamics. The

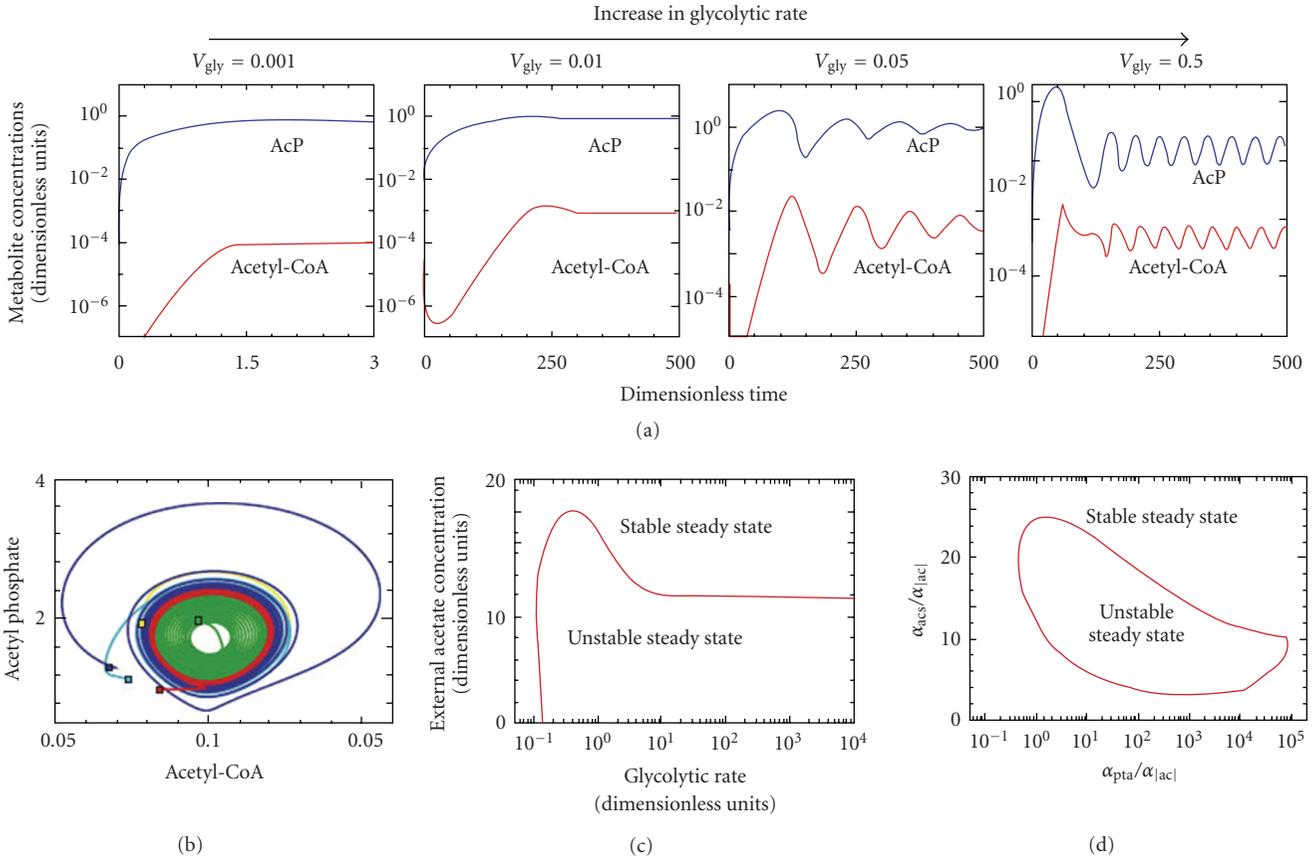


FIGURE 4: Sensitivity and bifurcation analyses of the metabolator model in Fung et al. [6]. (a) Sensitivity analysis: increase in glycolytic rate increases the oscillatory capability of the metabolator. The glycolytic rate  $V_{gly}$  is equal to 0.001, 0.01, 0.05, and 0.5 in the top four panels from left to right. (b) Phase plots obtained by perturbing the steady-state solution at  $V_{gly} = 1$  show that the oscillatory dynamics is limit cycle oscillation irrespective of the initial condition. The initial state of the oscillator is depicted with squares. (c) Hopf bifurcation analysis was used to construct a phase diagram of glycolytic rate versus external acetate concentration. The flux-sensitive nature of the oscillations is evident here; low glycolytic fluxes lead to a stable steady state with oscillations setting in beyond a threshold value of the glycolytic flux. (d) Another phase diagram suggests that at  $V_{gly} = 10$ , specific combinations of three protein levels are required to sustain oscillation. The variable  $\alpha_i$  represents rate of synthesis of protein  $i$  ( $i$ : LacI, Pta or Acs) (from Fung et al. [6], with permission).

use of bifurcation analysis was particularly useful as it facilitated the determination of the points at which the system transitions between stable and periodic states. We expect that as oscillator design develops [21], modeling will become ever more relevant in the design process.

### 8.2. Case Study II: Deterministic Model of a Synthetic Counter.

Another example of deterministic modeling is that of Friedland et al. [7], who developed synthetic counters in *E. coli* that can count up to two or three induction events. The first of these, a riboregulated transcriptional cascade (RTC) two-counter, has two nodes and is able to count up to two arabinose pulses by expressing a different protein in response to each pulse. This was extended to the RTC three-counter (Figure 5) which has three nodes and can count up to three pulses as follows. The constitutive promoter  $p_{Ltet0-1}$  drives transcription of T7 RNA polymerase (T7 RNAP), whose gene product binds the T7 promoter, which in turn drives the transcription of T3 RNAP. Similarly the protein of T3 RNAP binds the T3 promoter and ultimately drives

the transcription of GFP. All genes are further downregulated and upregulated by cis and trans elements of riboregulators. A cis repressor (cr) interacts with the downstream ribosome binding site (RBS) in such a manner as to prevent translation. The arabinose promoter  $p_{BAD}$  drives a transactivating, noncoding RNA (taRNA) that binds to cr in trans, thus relieving the inhibition of translation. Due to this design, protein synthesis at each node requires both transcription and translation to independently happen. Thus this cascade is able to count three arabinose pulses by expressing a different protein in response to each pulse.

The authors constructed and analyzed a mathematical model for the two- and three-counters. The model used equations of the form of (10) to describe the dynamics of the species in the circuit. The degradation terms in these equations were assumed to be simple exponential decays with a different rate constant for each species whereas the synthesis rates were rate laws that reflected how each species was synthesized. The Hill function was used to describe arabinose induction and the dynamics of GFP. Arabinose

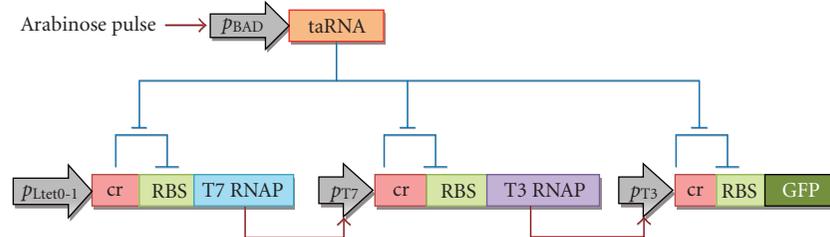


FIGURE 5: Design concepts for the RTC three-counter in Friedland et al. [7]. To count three induction events, the RTC three-counter employs a transcriptional cascade that has three nodes. The constitutive promoter  $p_{Ltet0-1}$  drives transcription of T7 RNA polymerase (T7 RNAP), whose gene product binds the T7 promoter, which in turn controls the transcription of T3 RNAP. Similarly the protein of T3 RNAP binds the T3 promoter and ultimately controls the transcription of GFP. All genes are further downregulated and upregulated by cis and trans elements of riboregulators. A cis repressor (cr) interacts with the downstream ribosome binding site (RBS) in such a manner as to prevent translation. The arabinose promoter  $p_{BAD}$  drives a transactivating, noncoding RNA (taRNA) that binds to cr in trans, thus relieving this inhibition of translation. Due to this design, protein synthesis at each node requires both transcription and translation to independently happen. Thus this cascade is able to count three arabinose pulses by expressing a different protein in response to each pulse (adapted from Friedland et al. [7]).

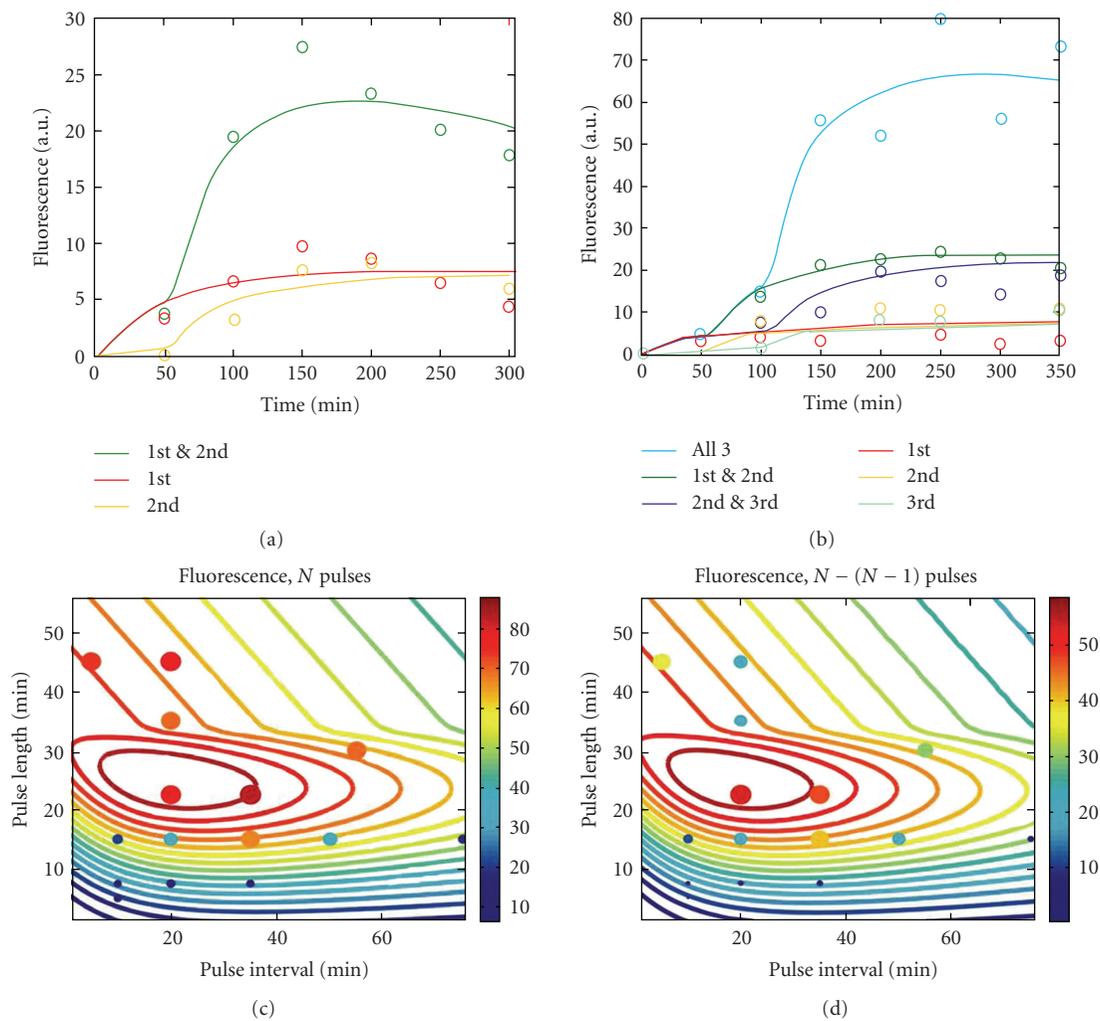


FIGURE 6: Modeling predictions and analysis of the RTC three-counter in Friedland et al. [7]. (a) A model of (a) the RTC two-counter and (b) the RTC three-counter with fitted parameters was simulated (solid line) and agrees well with experimental data (normalized fluorescence, solid dots). (c) The output of the RTC three-counter ( $N$  pulses) is simulated for a range of pulse lengths and intervals. The predictions (colored contour lines) match experimental results (solid circles), whose levels are indicated by both color and size. (d) Similar to (c), except that values shown here are the differences in the output of the three-counter after three ( $N$ ) pulses and two ( $N - 1$ ) pulses (from Friedland et al. [7], with permission).

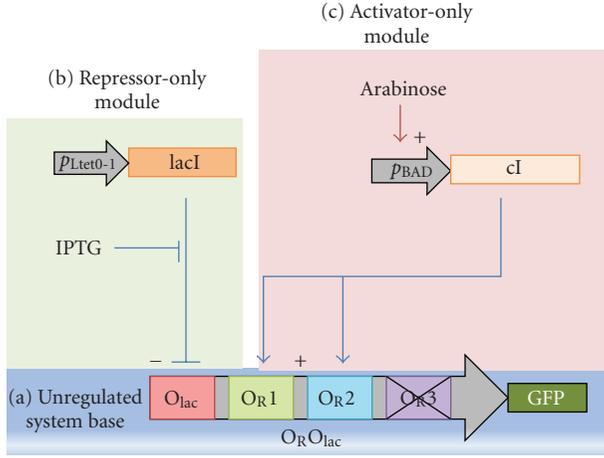


FIGURE 7: Design concept for the repressor-activator system in Guido et al. [8]. The repressor-activator system was assembled by combining (a) an unregulated system base with (b) a repressor-only module and (c) an activator-only module. The unregulated system consists of an  $O_R O_{lac}$  promoter driving GFP expression. The rectangular boxes in the promoter are the binding sites for the  $lacI$  and  $cI$  proteins. The cross at  $O_{R3}$  indicates a point mutation that inhibits  $cI$  binding at this site. In (b), the repressor-only module, the promoter  $p_{Ltet0-1}$  controls the transcription of  $lacI$ , which binds to the  $O_{lac}$  site to repress the  $O_R O_{lac}$  promoter. The extent of this effect can be tuned by using the  $lacI$  inhibitor IPTG, which reduces the binding of  $lacI$  protein to the  $O_{lac}$  site. In (c), the activator-only module, the  $p_{BAD}$  promoter controls the transcription of  $cI$ , which can bind to either the  $O_{R1}$  or the  $O_{R2}$  sites, sequentially or cooperatively, to activate the  $O_R O_{lac}$  promoter. The extent of this effect can be tuned through arabinose, which activates the  $p_{BAD}$  promoter (adapted from Guido et al. [8]).

pulse dynamics were modeled with two differential equations as follows. Arabinose consumption from the medium was modeled as a zero order rate law:

$$\frac{d[ara]}{dt} = -k_c, \quad (11)$$

whereas the decay of intracellular arabinose in cells suspended in arabinose-free media was modeled as a first order rate law:

$$\frac{d[ara]}{dt} = -k_d \cdot [ara]. \quad (12)$$

The arabinose pulses were mimicked by alternately using equation (11) and equation (12). The authors used optimization (implemented by a MATLAB routine *lsqcurvefit*) to evaluate (fit) the parameters in the model so as to agree with experimental data (Figures 6(a) and 6(b)). The model with fitted parameters was used to examine the effects of arabinose pulse frequency and length on the performance of the RTC three-counter (Figures 6(c) and 6(d)). The simulations indicated the ranges of pulse intervals (10 to 40 minutes) and pulse length (20 to 30 minutes) that would result in maximal system response (GFP fluorescence); these predictions were later confirmed by experiments (Figures 6(c) and 6(d)). The simulations indicated that the system

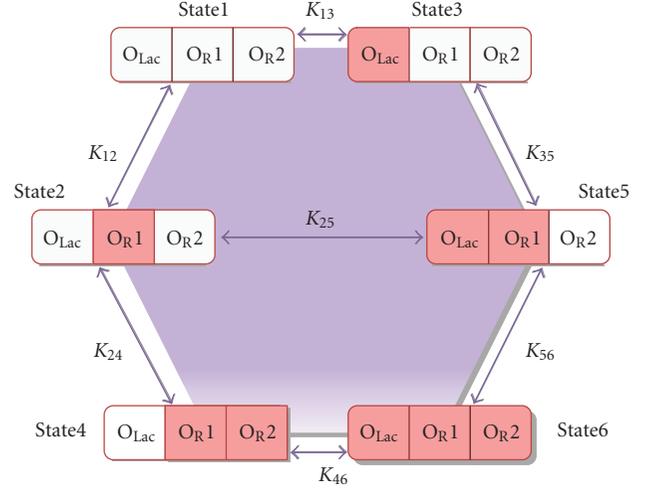


FIGURE 8: The six possible binding states in the repressor-activator model in Guido et al. [8]. The rectangular boxes represent the three available operator sites of engineered  $O_R O_{lac}$  promoter. Unbound sites are depicted with unfilled (white) boxes; filled (pink) boxes depict sites bound by the appropriate protein ( $lacI$  to  $O_{lac}$ ,  $cI$  to  $O_{R1}$  and  $O_{R2}$ ). Of the eight ( $2^3$ ) possible states, two are not feasible. This is because  $cI$  binds sequentially to the  $O_{R1}$  and  $O_{R2}$  sites so that  $O_{R2}$  cannot bind unless  $O_{R1}$  is also bound. The model parameters  $K_{12}$  through  $K_{56}$  are the rate constants for the transitions between each possible binding state (adapted from Guido et al. [8]).

response increased significantly when two or three pulses were delivered. Also indicated by the simulations was a small amount of leakage that occurred in response to partial pulses; this was also verified experimentally (Figures 6(a) and 6(b)). The authors further used this approach to construct a DNA invertase cascade (DIC) counter and discussed extending this counter with the use of other unique polymerases or recombinases to record  $N$  sequenced events.

The RTC two- and three-counters constitute an elegant example showing how synthetic circuit elements can be combined to recognize sequential events. Here too, the mathematical model was important in investigating the system dynamics and identifying the pulse length and interval that yielded the most effective response. Parameter estimation (and thereby optimization) provided crucial insights toward improving counter performance. As these counters are expanded to become capable of counting larger numbers of events (and thereby increasing in complexity), the role played by mathematical modeling in design will become increasingly important.

**8.3. Case Study III: Stochastic Model of a Bottom-Up Gene Regulatory Circuit.** Guido et al. [8] constructed regulatory networks by assembling simpler modules, such that the behavior of the network was predictable from that of the components. The authors engineered the  $O_R O_{lac}$  promoter such that it caused both repression and activation of gene expression in *E. coli*. For this, they combined an unregulated promoter, a repressor-only and an activator-only system (Figure 7). The unregulated system (Figure 7(a)) consists

**Steps in the model of Guido et al. [8]**

**Step 1.** Calculating the state probability of each promoter  
Using  $p_1$  as the reference, the probabilities for other five states could be described as:

$$p_2 = k_{12}p_1$$

$$p_3 = k_{13}p_1$$

$$p_4 = k_{12}k_{24}p_1$$

$$p_5 = k_{13}k_{35}p_1$$

$$p_6 = k_{12}k_{24}k_{46}p_1$$

From the first constraint,  $p_1$ , the probability of the state of the unregulated promoter, is obtained as:

$$p_1 = \frac{1}{1 + k_{12} + k_{13} + k_{12}k_{24} + k_{13}k_{35} + k_{12}k_{24}k_{46}}$$

**Probability constraint:**

$$\sum_{i=1}^6 p_i = 1$$

**Thermodynamic equilibrium constraints:**

$$\frac{k_{12}k_{25}}{k_{13}k_{35}} = 1, \frac{k_{24}k_{46}}{k_{25}k_{56}} = 1$$

**Step 2.** Estimating parameters  
Factors controlling equilibrium constants:  $k_{13}, k_{25}$  and  $k_{46}$  depend on the concentration of IPTG  
 $k_{12}, k_{24}, k_{35}$  and  $k_{56}$  depend on the concentration of arabinose

For the repressor-only system:  
 $k_{12} = k_{24} = k_{35} = k_{56} = 0$   
For activator-only system:  
 $k_{13} = k_{25} = k_{46} = 0$

**Step 3.** The model response (average mRNA synthesis rate)  
The average synthesis rate is given by:  
 $\bar{y}_m = p_1 + g_2 p_2 + g_3 p_3 + g_4 p_4 + g_5 p_5 + g_6 p_6$   
where  $g_i$  is a nonlinear function that represents the mRNA synthesis rate in the  $i$ th promoter state.

**Step 4.** Extend the model with stochastic effects  
Incorporate cell growth, synthesis and degradation rates of mRNA and GFP, multimerization of cI.

FIGURE 9: Steps in the repressor-activator mathematical model in Guido et al. [8].

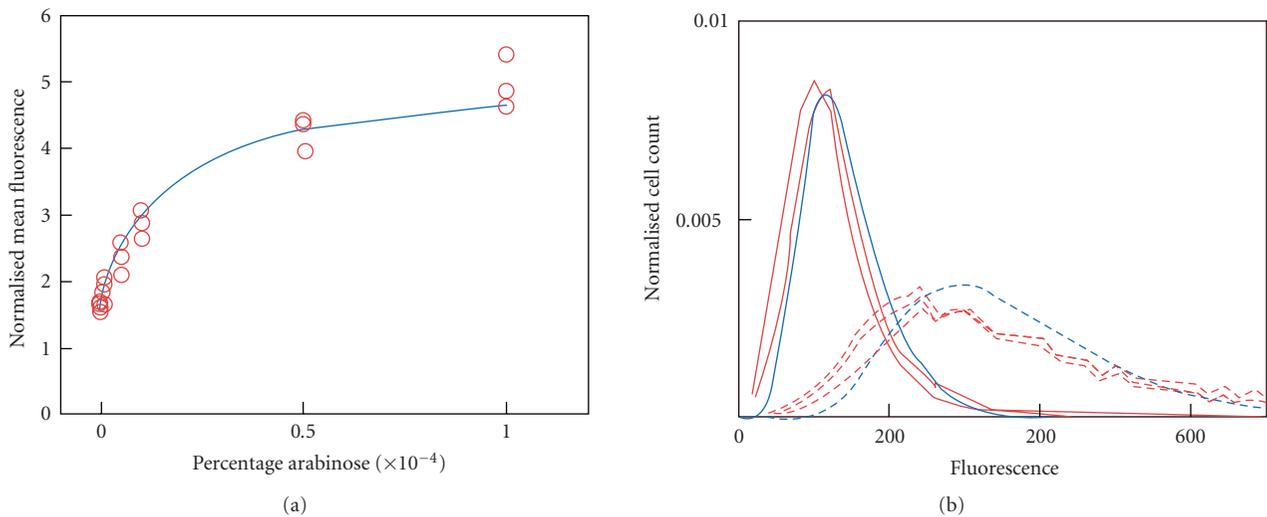


FIGURE 10: Model predictions and analysis of the repressor-activator system in Guido et al. [8]. (a) Simulations with baseline deterministic model: dependence of the model response (GFP fluorescence) of the repressor-activator system on arabinose concentration, with IPTG concentration maintained constant. The model simulation is shown with a blue line and experimental data points are shown with red circles. (b) Simulations with the extended stochastic model: distribution of GFP fluorescence within the cell population for levels of arabinose ranging from no arabinose (solid lines toward left) to the maximum possible level of arabinose (dashed lines toward right); IPTG concentration is maintained constant throughout. Stochastic model simulations are shown with blue lines and experimental data is shown with red lines (from Guido et al. [8], with permission).

of an  $O_R O_{lac}$  promoter driving GFP expression. In the repressor-only module (Figure 7(b)), the promoter  $p_{Ltet0-1}$  controls the transcription of lacI, which binds to the  $O_{lac}$  site to repress the  $O_R O_{lac}$  promoter. This binding can be tuned by with the lacI inhibitor isopropyl-b-D-thiogalactopyranoside

(IPTG), which reduces the binding of lacI protein to the  $O_{lac}$  site. In the activator-only module (Figure 7(c)), the  $p_{BAD}$  promoter controls the transcription of cI, which can bind sequentially or cooperatively to either the  $O_{R1}$  or the  $O_{R2}$  sites of the  $O_R O_{lac}$  promoter to activate it. This effect is

tunable with arabinose, which activates the  $p_{BAD}$  promoter. The repressor-activator system is constructed by combining all three of these modules.

The authors first developed a baseline deterministic model for the repressor-activator system and then extended it to include stochastic effects. A quasi-equilibrium state was assumed for the promoter  $O_{R}O_{lac}$  such that there were six possible chemical states of the promoter (Figure 8). The interconversions between the states were modeled with the equation:

$$K_{ij} = K_{ij}^{eq}[P], \quad (13)$$

where  $K_{ij}$  is the transition rate from promoter state  $S_i$  to  $S_j$ ,  $K_{ij}^{eq}$  is the equilibrium constant for the interconversion, and  $[P]$ , depending on the reaction, is the concentration of either  $lacI$  or  $cl$ . The probabilities of each state were functions of the transition rates  $K_{ij}$  and the average GFP synthesis rate was in turn a function of the state probabilities (see discussion around equation (4) and Figure 9 for details). The authors extended the model with stochastic effects so that it was able to simulate cell growth and the distribution of GFP within the cell population. For example cell growth was modeled by treating the cell volume as a random variable with an exponential distribution. The model assumed that at every instance of cell division, the volume halves and the cellular components (mRNA and GFP) are distributed amongst the daughter cells in a binomial distribution. The simulations of the deterministic and stochastic parts of the model are shown in Figure 10. The parameters in this model were determined by a gradient search local optimization algorithm and all stochastic modeling was implemented through the Gillespie Monte Carlo method realized by BioNetS software. The authors experimentally measured all mRNA synthesis rates and GFP fluorescence relative to the unregulated system and chose the parameter values that best fitted the fluorescence results shown in Figure 10. The model slightly underestimates the experimentally observed variability of fluorescence within the cell population which indicates that effects not considered in the model might be occurring.

To further verify the predictive power of the model and test its stochastic aspects, the authors expanded the circuit to include a positive feedback. This was accomplished adding the  $cl$  gene in front of GFP so that the repressor-activator system transcribes both  $cl$  and GFP simultaneously. This amounts to positive feedback because the product of the  $cl$  gene activates gene expression in the system. To effectively model this feedback loop, the model was extended to include synthesis, degradation, and multimerization of  $cl$ . This model with positive feedback exhibited very good agreement with experimental data. This agreement validated the bottom-up approach employed by the authors to study regulatory systems. Furthermore, the model was used to make the counterintuitive prediction that cessation of cell growth and division increases noise in protein expression levels, which was also verified experimentally.

Although other researchers similarly built larger regulatory systems from simple ones (e.g., [89]), the work by Guido

et al. emphasized the important role played by mathematical modeling in the design and in making counterintuitive predictions about system dynamics. We expect that in the future, mathematical modeling along with transcription network analyses will become indispensable in the design of more complex regulatory networks.

## 9. Conclusion and Outlook

Mathematical modeling is an often indispensable tool in synthetic biology. The mathematical techniques of parameter estimation as well as sensitivity and bifurcation analyses can be crucial to the development of a model intended to mimic a complex system. Modeling also plays an important role in phenotypic analyses such as metabolic flux analysis or transcription network analysis.

A mathematical model is akin to a road map that provides a visualization of a geographical area. Although the map may not describe every detail of the landscape, it contains adequate information to enable users to plan a journey; a mathematical model is similar in scope [4]. A seminal review on mathematical modeling [90] stated that the purpose of models is to discern which parts and connections of a system are significant, to unravel new strategies, or sometimes to correct conventional wisdom. Examples of these uses abound in synthetic biology, where models have been employed to identify which regions in parameter space cause a system to behave in a desired manner (e.g., [6] and case study I) or what parameter values result in the most effective design (e.g., [7] and case study II). Furthermore, models have also been employed to understand the global dynamics of a system from known behaviors of its component units and have made counterintuitive predictions that were later verified experimentally (e.g., [8] and case study III). We anticipate that as experimental advances in synthetic biology produce increasingly complex circuits, mathematical modeling will play an ever more important function as a bridge between concept and realization.

## Acknowledgments

This work was funded by the Department of Chemical and Biomolecular Engineering, University of Maryland (faculty startup grant to GS), the A. James Clark School of Engineering, University of Maryland (Minta Martin award to GS), and Maryland Industrial Partnerships (MIPS) (award no. 4426). Y. Zheng was cofunded by a Jan and Anneke Sengers fellowship.

## References

- [1] F. J. Isaacs, D. J. Dwyer, and J. J. Collins, "RNA synthetic biology," *Nature Biotechnology*, vol. 24, no. 5, pp. 545–554, 2006.
- [2] M. Heinemann and S. Panke, "Synthetic biology—putting engineering into biology," *Bioinformatics*, vol. 22, no. 22, pp. 2790–2799, 2006.
- [3] E. Andrianantoandro, S. Basu, D. K. Karig, and R. Weiss, "Synthetic biology: new engineering rules for an emerging

- discipline,” *Molecular Systems Biology*, vol. 2, Article ID 2006.0028, 2006.
- [4] E. L. Haseltine and F. H. Arnold, “Synthetic gene circuits: design with directed evolution,” *Annual Review of Biophysics and Biomolecular Structure*, vol. 36, pp. 1–19, 2007.
- [5] D. Endy, “Foundations for engineering biology,” *Nature*, vol. 438, no. 7067, pp. 449–453, 2005.
- [6] E. Fung, W. W. Wong, J. K. Suen, T. Bulter, S.-G. Lee, and J. C. Liao, “A synthetic gene-metabolic oscillator,” *Nature*, vol. 435, no. 7038, pp. 118–122, 2005.
- [7] A. E. Friedland, T. K. Lu, X. Wang, D. Shi, G. Church, and J. J. Collins, “Synthetic gene networks that count,” *Science*, vol. 324, no. 5931, pp. 1199–1202, 2009.
- [8] N. J. Guido, X. Wang, D. Adalsteinsson, et al., “A bottom-up approach to gene regulation,” *Nature*, vol. 439, no. 7078, pp. 856–860, 2006.
- [9] D. Chandran, W. B. Copeland, S. C. Sleight, and H. M. Sauro, “Mathematical modeling and synthetic biology,” *Drug Discovery Today: Disease Models*, vol. 5, no. 4, pp. 299–309, 2008.
- [10] Y. N. Kaznessis, “Models for synthetic biology,” *BMC Systems Biology*, vol. 1, article 47, 2007.
- [11] O. Brandman, J. E. Ferrell Jr., R. Li, and T. Meyer, “Systems biology: interlinked fast and slow positive feedback loops drive reliable cell decisions,” *Science*, vol. 310, no. 5747, pp. 496–498, 2005.
- [12] A. Ma’ayan, S. L. Jenkins, S. Neves, et al., “Formation of regulatory patterns during signal propagation in a mammalian cellular network,” *Science*, vol. 309, no. 5737, pp. 1078–1083, 2005.
- [13] J. E. Lunn, “Compartmentation in plant metabolism,” *Journal of Experimental Botany*, vol. 58, no. 1, pp. 35–47, 2007.
- [14] B. N. Kholodenko, “Cell-signalling dynamics in time and space,” *Nature Reviews Molecular Cell Biology*, vol. 7, no. 3, pp. 165–176, 2006.
- [15] M. J. Hounslow, R. L. Ryall, and V. R. Marshall, “A discretized population balance for nucleation, growth, and aggregation,” *AIChE Journal*, vol. 34, no. 11, pp. 1821–1832, 1988.
- [16] A. Ma’ayan, “Insights into the organization of biochemical regulatory networks using graph theory analyses,” *Journal of Biological Chemistry*, vol. 284, no. 9, pp. 5451–5455, 2009.
- [17] S. Mangan and U. Alon, “Structure and function of the feed-forward loop network motif,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 21, pp. 11980–11985, 2003.
- [18] J. E. Ferrell Jr., “Self-perpetuating states in signal transduction: positive feedback, double-negative feedback and bistability,” *Current Opinion in Cell Biology*, vol. 14, no. 2, pp. 140–148, 2002.
- [19] U. S. Bhalla and R. Iyengar, “Emergent properties of networks of biological signaling pathways,” *Science*, vol. 283, no. 5400, pp. 381–387, 1999.
- [20] A. Becskei and L. Serrano, “Engineering stability in gene networks by autoregulation,” *Nature*, vol. 405, no. 6786, pp. 590–593, 2000.
- [21] J. Stricker, S. Cookson, M. R. Bennett, W. H. Mather, L. S. Tsimring, and J. Hasty, “A fast, robust and tunable synthetic gene oscillator,” *Nature*, vol. 456, no. 7221, pp. 516–519, 2008.
- [22] M. B. Elowitz and S. Leibler, “A synthetic oscillatory network of transcriptional regulators,” *Nature*, vol. 403, no. 6767, pp. 335–338, 2000.
- [23] J. J. Tyson, K. Chen, and B. Novak, “Network dynamics and cell physiology,” *Nature Reviews Molecular Cell Biology*, vol. 2, no. 12, pp. 908–916, 2001.
- [24] L. Allen, *An Introduction to Stochastic Processes with Applications to Biology*, Pearson Education, Upper Saddle River, NJ, USA, 2003.
- [25] H. Salis and Y. N. Kaznessis, “An equation-free probabilistic steady-state approximation: dynamic application to the stochastic simulation of biochemical reaction networks,” *Journal of Chemical Physics*, vol. 123, no. 21, Article ID 214106, 16 pages, 2005.
- [26] H. Salis and Y. Kaznessis, “Accurate hybrid stochastic simulation of a system of coupled chemical or biochemical reactions,” *Journal of Chemical Physics*, vol. 122, no. 5, Article ID 054103, 13 pages, 2005.
- [27] L. M. Tuttle, H. Salis, J. Tomshine, and Y. N. Kaznessis, “Model-driven designs of an oscillating gene network,” *Biophysical Journal*, vol. 89, no. 6, pp. 3873–3883, 2005.
- [28] B. Zhou, D. Beckwith, L. R. Jarboe, and J. C. Liao, “Markov chain modeling of pyelonephritis-associated pili expression in uropathogenic *Escherichia coli*,” *Biophysical Journal*, vol. 88, no. 4, pp. 2541–2553, 2005.
- [29] L. R. Jarboe, D. Beckwith, and J. C. Liao, “Stochastic modeling of the phase-variable pap operon regulation in uropathogenic *Escherichia coli*,” *Biotechnology and Bioengineering*, vol. 88, no. 2, pp. 189–203, 2004.
- [30] M. Sugimoto, S. Kikuchi, and M. Tomita, “Reverse engineering of biochemical equations from time-course data by means of genetic programming,” *BioSystems*, vol. 80, no. 2, pp. 155–164, 2005.
- [31] L. M. Tran, M. L. Rizk, and J. C. Liao, “Ensemble modeling of metabolic networks,” *Biophysical Journal*, vol. 95, no. 12, pp. 5606–5617, 2008.
- [32] E. O. Voit, *Computational Analysis of Biochemical Systems*, Cambridge University Press, Cambridge, UK, 2000.
- [33] M. Rodriguez-Fernandez, P. Mendes, and J. R. Banga, “A hybrid approach for efficient and robust parameter estimation in biochemical pathways,” *BioSystems*, vol. 83, no. 2-3, pp. 248–265, 2006.
- [34] E. Walter and L. Pronzato, *Identification of Parameter Models from Experimental Data*, Springer, Berlin, Germany, 1997.
- [35] L. Ljung, *System Identification: Theory for the User*, Prentice Hall, Englewood Cliffs, NJ, USA, 1999.
- [36] W. E. Stewart, M. Caracotsios, and J. P. Sørensen, “Parameter estimation from multiresponse data,” *AIChE Journal*, vol. 38, no. 5, pp. 641–650, 1992.
- [37] D. Ropers, H. de Jong, M. Page, D. Schneider, and J. Geiselmann, “Qualitative simulation of the carbon starvation response in *Escherichia coli*,” *BioSystems*, vol. 84, no. 2, pp. 124–152, 2006.
- [38] E. M. Ozbudak, M. Thattai, I. Kurtser, A. D. Grossman, and A. van Oudenaarden, “Regulation of noise in the expression of a single gene,” *Nature Genetics*, vol. 31, no. 1, pp. 69–73, 2002.
- [39] J. Paulsson, “Models of stochastic gene expression,” *Physics of Life Reviews*, vol. 2, no. 2, pp. 157–175, 2005.
- [40] M. B. Elowitz, A. J. Levine, E. D. Siggia, and P. S. Swain, “Stochastic gene expression in a single cell,” *Science*, vol. 297, no. 5584, pp. 1183–1186, 2002.
- [41] I. Swameye, T. G. Müller, J. Timmer, O. Sandra, and U. Klingmüller, “Identification of nucleocytoplasmic cycling as a remote sensor in cellular signaling by databased modeling,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 3, pp. 1028–1033, 2003.
- [42] J. Timmer, T. G. Müller, I. Swameye, O. Sandra, and U. Klingmüller, “Modeling the nonlinear dynamics of cellular

- signal transduction,” *International Journal of Bifurcation & Chaos in Applied Sciences & Engineering*, vol. 14, no. 6, pp. 2069–2079, 2004.
- [43] S. Audoly, L. D’Angiò, M. P. Saccomani, and C. Cobelli, “Global identifiability of linear compartmental models—a computer algebra algorithm,” *IEEE Transactions on Biomedical Engineering*, vol. 45, no. 1, pp. 36–47, 1998.
- [44] E. Cinquemani, A. Miliadis-Argeitis, S. Summers, and J. Lygeros, “Stochastic dynamics of genetic networks: modelling and parameter identification,” *Bioinformatics*, vol. 24, no. 23, pp. 2748–2754, 2008.
- [45] P. Mendes and D. B. Kell, “Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation,” *Bioinformatics*, vol. 14, no. 10, pp. 869–883, 1998.
- [46] C. G. Moles, P. Mendes, and J. R. Banga, “Parameter estimation in biochemical pathways: a comparison of global optimization methods,” *Genome Research*, vol. 13, no. 11, pp. 2467–2474, 2003.
- [47] G. R. Wood, D. L. J. Alexander, and D. W. Bulger, “Approximation of the distribution of convergence times for stochastic global optimisation,” *Journal of Global Optimization*, vol. 22, no. 1–4, pp. 271–284, 2002.
- [48] E. Onbaşoğlu and L. Özdamar, “Parallel simulated annealing algorithms in global optimization,” *Journal of Global Optimization*, vol. 19, no. 1, pp. 27–50, 2001.
- [49] E. Balsa-Canto, M. Rodriguez-Fernandez, and J. R. Banga, “Optimal design of dynamic experiments for improved estimation of kinetic parameters of thermal degradation,” *Journal of Food Engineering*, vol. 82, no. 2, pp. 178–188, 2007.
- [50] M. Garey and D. Johnson, *Computers and Intractability: A Guide to the Theory of NP-Completeness*, Freeman, San Francisco, Calif, USA, 1979.
- [51] K.-H. Cho, S.-Y. Shin, W. Kolch, and O. Wolkenhauer, “Experimental design in systems biology, based on parameter sensitivity analysis using a monte carlo method: a case study for the TNF $\alpha$ -mediated NF- $\kappa$ B signal transduction pathway,” *Simulation*, vol. 79, no. 12, pp. 726–739, 2003.
- [52] X.-J. Feng, S. Hooshangi, D. Chen, G. Li, R. Weiss, and H. Rabitz, “Optimizing genetic circuits by global sensitivity analysis,” *Biophysical Journal*, vol. 87, no. 4, pp. 2195–2202, 2004.
- [53] J. D. Keasling, “Synthetic biology for synthetic chemistry,” *ACS Chemical Biology*, vol. 3, no. 1, pp. 64–76, 2008.
- [54] C. Tan, H. Song, J. Niemi, and L. You, “A synthetic biology challenge: making cells compute,” *Molecular BioSystems*, vol. 3, no. 5, pp. 343–353, 2007.
- [55] V. Chickarmane, B. N. Kholodenko, and H. M. Sauro, “Oscillatory dynamics arising from competitive inhibition and multisite phosphorylation,” *Journal of Theoretical Biology*, vol. 244, no. 1, pp. 68–76, 2007.
- [56] T. S. Gardner, C. R. Cantor, and J. J. Collins, “Construction of a genetic toggle switch in *Escherichia coli*,” *Nature*, vol. 403, no. 6767, pp. 339–342, 2000.
- [57] J. Garcia-Ojalvo, M. B. Elowitz, and S. H. Strogatz, “Modeling a synthetic multicellular clock: repressilators coupled by quorum sensing,” *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 30, pp. 10955–10960, 2004.
- [58] J. Hasty, M. Dolnik, V. Rottschäfer, and J. J. Collins, “Synthetic gene network for entraining and amplifying cellular oscillations,” *Physical Review Letters*, vol. 88, no. 14, Article ID 148101, 4 pages, 2002.
- [59] J. Noirel, S. Y. Ow, G. Sanguinetti, and P. C. Wright, “Systems biology meets synthetic biology: a case study of the metabolic effects of synthetic rewiring,” *Molecular BioSystems*, vol. 5, no. 10, pp. 1214–1223, 2009.
- [60] V. Hatzimanikatis, C. A. Floudas, and J. E. Bailey, “Analysis and design of metabolic reaction networks via mixed-integer linear optimization,” *AIChE Journal*, vol. 42, no. 5, pp. 1277–1292, 1996.
- [61] G. Sriram, D. B. Fulton, V. V. Iyer, et al., “Quantification of compartmented metabolic fluxes in developing soybean embryos by employing biosynthetically directed fractional  $^{13}\text{C}$  labeling, two-dimensional [ $^{13}\text{C}$ ,  $^1\text{H}$ ] nuclear magnetic resonance, and comprehensive isotopomer balancing,” *Plant Physiology*, vol. 136, no. 2, pp. 3043–3057, 2004.
- [62] K. Nöh and W. Wiechert, “Experimental design principles for isotopically instationary  $^{13}\text{C}$  labeling experiments,” *Biotechnology and Bioengineering*, vol. 94, no. 2, pp. 234–251, 2006.
- [63] M. R. Antoniewicz, J. K. Kelleher, and G. Stephanopoulos, “Elementary metabolite units (EMU): a novel framework for modeling isotopic distributions,” *Metabolic Engineering*, vol. 9, no. 1, pp. 68–86, 2007.
- [64] W. Wiechert, M. Möllney, N. Isermann, M. Wurzel, and A. A. De Graaf, “Bidirectional reaction steps in metabolic networks: III. Explicit solution and analysis of isotopomer labeling systems,” *Biotechnology and Bioengineering*, vol. 66, no. 2, pp. 69–85, 1999.
- [65] G. Sriram and J. V. Shanks, “Improvements in metabolic flux analysis using carbon bond labeling experiments: bondomer balancing and Boolean function mapping,” *Metabolic Engineering*, vol. 6, no. 2, pp. 116–132, 2004.
- [66] W. A. van Winden, J. J. Heijnen, and P. J. T. Verheijen, “Cumulative bondomers: a new concept in flux analysis from 2D [ $^{13}\text{C}$ ,  $^1\text{H}$ ] COSY NMR data,” *Biotechnology and Bioengineering*, vol. 80, no. 7, pp. 731–745, 2002.
- [67] J. D. Young, J. L. Walther, M. R. Antoniewicz, H. Yoo, and G. Stephanopoulos, “An elementary metabolite unit (EMU) based method of isotopically nonstationary flux analysis,” *Biotechnology and Bioengineering*, vol. 99, no. 3, pp. 686–699, 2008.
- [68] J. M. Lee, E. P. Gianchandani, and J. A. Papin, “Flux balance analysis in the era of metabolomics,” *Briefings in Bioinformatics*, vol. 7, no. 2, pp. 140–150, 2006.
- [69] J. S. Edwards and B. O. Palsson, “Metabolic flux balance analysis and the in silico analysis of *Escherichia coli* K-12 gene deletions,” *BMC Bioinformatics*, vol. 1, article 1, 2000.
- [70] K. J. Kauffman, P. Prakash, and J. S. Edwards, “Advances in flux balance analysis,” *Current Opinion in Biotechnology*, vol. 14, no. 5, pp. 491–496, 2003.
- [71] O. Resendis-Antonio, J. L. Reed, S. Encarnación, J. Collado-Vides, and B. Ø. Palsson, “Metabolic reconstruction and modeling of nitrogen fixation in *Rhizobium etli*,” *PLoS Computational Biology*, vol. 3, no. 10, article e192, 2007.
- [72] M. AbuOun, P. F. Suthers, G. I. Jones, et al., “Genome scale reconstruction of a salmonella metabolic model: comparison of similarity and differences with a commensal *Escherichia coli* strain,” *Journal of Biological Chemistry*, vol. 284, no. 43, pp. 29480–29488, 2009.
- [73] A. M. Feist, C. S. Henry, J. L. Reed, et al., “A genome-scale metabolic reconstruction for *Escherichia coli* K-12 MG1655 that accounts for 1260 ORFs and thermodynamic information,” *Molecular Systems Biology*, vol. 3, article 121, 2007.
- [74] P. Geigenberger, M. Stitt, and A. R. Fernie, “Metabolic control analysis and regulation of the conversion of sucrose to starch

- in growing potato tubers," *Plant, Cell & Environment*, vol. 27, no. 6, pp. 655–673, 2004.
- [75] U. S. Ramli, J. J. Salas, P. A. Quant, and J. L. Harwood, "Metabolic control analysis reveals an important role for diacylglycerol acyltransferase in olive but not in oil palm lipid accumulation," *FEBS Journal*, vol. 272, no. 22, pp. 5764–5770, 2005.
- [76] B. M. Bakker, H. V. Westerhoff, F. R. Opperdoes, and P. A. M. Michels, "Metabolic control analysis of glycolysis in trypanosomes as an approach to improve selectivity and effectiveness of drugs," *Molecular and Biochemical Parasitology*, vol. 106, no. 1, pp. 1–10, 2000.
- [77] D. A. Fell, "Metabolic control analysis: a survey of its theoretical and experimental development," *Biochemical Journal*, vol. 286, no. 2, pp. 313–330, 1992.
- [78] D. Fell, "Metabolic control analysis," in *Systems Biology*, L. Alberghina and H. Westerhoff, Eds., pp. 69–80, Springer, Berlin, Germany, 2005.
- [79] C. C. Guet, M. B. Elowitz, W. Hsing, and S. Leibler, "Combinatorial synthesis of genetic networks," *Science*, vol. 296, no. 5572, pp. 1466–1470, 2002.
- [80] J. C. Liao, R. Boscolo, Y.-L. Yang, L. M. Tran, C. Sabatti, and V. P. Roychowdhury, "Network component analysis: reconstruction of regulatory signals in biological systems," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 26, pp. 15522–15527, 2003.
- [81] Z. Li and C. Chan, "Extracting novel information from gene expression data," *Trends in Biotechnology*, vol. 22, no. 8, pp. 381–383, 2004.
- [82] S. Raychaudhuri, J. M. Stuart, and R. B. Altman, "Principal components analysis to summarize microarray experiments: application to sporulation time series," in *Pacific Symposium on Biocomputing*, pp. 455–466, Island of Oahu, Hawaii, USA, 2000.
- [83] M. K. S. Yeung, J. Tegnér, and J. J. Collins, "Reverse engineering gene networks using singular value decomposition and robust regression," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 99, no. 9, pp. 6163–6168, 2002.
- [84] N. S. Holter, A. Maritan, M. Cieplak, N. V. Fedoroff, and J. R. Banavar, "Dynamic modeling of gene expression data," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 98, no. 4, pp. 1693–1698, 2001.
- [85] W. Liebermeister, "Linear modes of gene expression determined by independent component analysis," *Bioinformatics*, vol. 18, no. 1, pp. 51–60, 2002.
- [86] Z. Li, S. M. Shaw, M. J. Yedwabnick, and C. Chan, "Using a state-space model with hidden variables to infer transcription factor activities," *Bioinformatics*, vol. 22, no. 6, pp. 747–754, 2006.
- [87] C. T. Harbison, D. B. Gordon, T. I. Lee, et al., "Transcriptional regulatory code of a eukaryotic genome," *Nature*, vol. 431, no. 7004, pp. 99–104, 2004.
- [88] M. Hucka, A. Finney, H. M. Sauro, et al., "The systems biology markup language (SBML): a medium for representation and exchange of biochemical network models," *Bioinformatics*, vol. 19, no. 4, pp. 524–531, 2003.
- [89] S. Bornholdt, "Less is more in modeling large genetic networks," *Science*, vol. 310, no. 5747, pp. 449–451, 2005.
- [90] J. E. Bailey, "Mathematical modeling and analysis in biochemical engineering: past accomplishments and future opportunities," *Biotechnology Progress*, vol. 14, no. 1, pp. 8–20, 1998.



**Hindawi**

Submit your manuscripts at  
<http://www.hindawi.com>

