A SYSTEMS APPROACH TO MODELING AND ANALYZING BIOLOGICAL REGULATION

Francis J. Doyle III *,1

* *Department of Chemical Engineering, University of California, Santa Barbara, USA*

Abstract: Understanding regulation is a critical hurdle in unraveling complex biological systems. As gene-level architectures become known, the open challenge is to assign predictable behavior to a known gene structure, the so-called "genotype-to-phenotype" problem. In response to this challenge, the discipline of systems biology has emerged with an integrative perspective towards determining complex systems behavior. This research area lies at the intersection of classical fields such as microbiology and systems engineering, with strong influences from the more recent fields of informatics and genomics. In this paper, an overview of a number of quantitative tools from systems theory will be presented as enabling methodologies for unraveling biological regulatory systems, with an emphasis on (i) sensitivity analysis, (ii) identification methods, and (iii) optimization approaches.

Keywords: Systems biology, regulation, gene networks, circadian rhythm

1. INTRODUCTION

The term "complexity" is often invoked in the description of biophysical networks that underlie gene regulation in biological organisms. There are categorically two distinct characterizations of complexity: (i) the classical notion of behavior associated with the mathematical properties of chaos and bifurcations, and (ii) the descriptive or topological notion of a large number of constitutive elements with nontrivial connectivity. In both biological and more general contexts, a key implication of complexity is that the underlying system is difficult to understand and to verify (Wen *et al.*, 1998). Simple low-order mathematical models can be constructed that yield chaotic behavior on one hand, and yet rich complex biophysical networks may be designed to reinforce reliable execution of simple tasks or behaviors (Lauffenburger, 2000).

A systematic approach to analyzing complexity in biophysical networks was previously untenable, owing to the lack of suitable measurements and also the limitations imposed in simulating complex mathematical models. Advances in molecular biology over the past decade have made it possible to probe experimentally the causal relationships between microbiological processes initiated by individual molecules within a cell, and their macroscopic phenotypic effects on cells and organisms. These studies provide increasingly detailed insights into the underlying networks, circuits, and pathways responsible for the basic functionality and robustness of biological systems and create new and exciting opportunities for the development of quantitative and predictive modeling and simulation tools. Model development involves translating identified biological processes into coupled dynamical equations which are amenable to numerical simulation and analysis. These equations describe the interactions between various constituents and the environment, and involve

[1] e-mail: doyle@engineering.ucsb.edu

multiple feedback loops, responsible for system regulation, and noise attenuation and amplification.

The discipline of *Systems Biology* has emerged in response to these challenges (Kitano, 2002), and combines approaches and methods from systems engineering, computational biology, statistics, genomics, molecular biology, biophysics, and other fields. The recurring themes include: (i) integrative viewpoints towards unraveling complex dynamical systems, and (ii) tight iterations between experiments, modeling, and hypothesis generation.

The perspective adopted in this paper is that systems engineering tools, including model identification; sensitivity analysis, and dynamic programming, find unique roles in characterizing the rich behavior exhibited by such biological systems. The systems perspective is also valuable in analyzing the integrative behavior of such complex multiscale stochastic systems, as opposed to traditional reduction techniques.

2. ELEMENTS OF SYSTEMS BIOLOGY

In this section, three topics in systems biology will be described that have relevance for systems engineering tools and methodologies. They include: (i) the recurrence of elementary motifs in biological networks, (ii) a brief classification of computational modeling approaches to biological regulation, and (iii) the detailed treatment of feedback control structures in some specific biological systems.

2.1 Motifs in Regulation

The biophysical networks under consideration can be decomposed into modular components that recur across and within given organisms. One hierarchical classification is to label the top level as a *network*, which is comprised of interacting regulatory *motifs* consisting of groups of 2-4 genes (Lee *et al.*, 2002; Shen-Orr *et al.*, 2002; Zak *et al.*, 2003*a*). At the lowest level in this hierarchy is the *module* that describes transcriptional regulation, of which a nice example is given in (Barkai and Leibler, 2000).

At the *motif* level, one can use pattern searching techniques to determine the frequency of occurrence of these simple motifs (Shen-Orr *et al.*, 2002), leading to the postulation that these are basic building blocks in biological networks. Of relevance to the present discussion is the fact that many of these components have direct analogs in system engineering architectures. Consider the three dominant network motifs found in *E. coli* (Shen-Orr *et al.*, 2002):

- **feedforward loop** - one transcription factor regulates another factor, and in turn the pair jointly regulate a third transcript factor
- **single input module (SIM)** - in control terminology, a single-input multiple output (SIMO) block architecture
- **densely overlapping regulons (DOR)** - in control terminology, a single-input multiple output (SIMO) block architecture

Similar studies in a completely different organism, *S. cerevisiae*, yielded six related or overlapping network motifs (Lee *et al.*, 2002):

- **autoregulatory motif** - in which a regulator binds to the promotor region of its own gene
- **feedforward loop** - as noted above
- **multi-component loop** - effectively, a closed-loop with two or more transcription factors
- **regulator chain** - cascade of serial transcription factor interactions
- **single input module** - as noted above (SIM)
- **multi-input module** - natural extension of preceding motif

In effect, these studies prove that, in both eukaryotic and prokaryotic systems, cell function is controlled by sophisticated networks of control loops which are cascading onto, and interconnected with, other (transcriptional) control loops. The noteworthy insight here is that the complex networks that underly biological regulation appear to be constructed of elementary *systems* components, not unlike a digital circuit. This lends credibility to the notion that analysis tools from systems engineering should find relevance in this problem domain.

Reverse engineering, or the process of unraveling the functionality of such motifs is one of the major goals of systems biology, but *forward engineering*, in effect the dual problem, is also of considerable interest and is concerned with the active design of biological networks, typically for biomedical or biotechnological applications. In that direction, early work has shown great promise for the development of synthetic gene networks, using basic building blocks to achieve a desired function (for example, (Judd *et al.*, 2000)). A novel 3-node oscillating circuit, dubbed the "repressilator", was studied using flow cytometry imaging techniques (Elowitz and Leibler, 2000). Several other examples of synthetic functions (switches and oscillators) have been realized (Hasty *et al.*, 2001*b*).

2.2 Dynamic Models of Gene Regulation

In this section, three general classes of modeling techniques that have been applied to gene regulation are reviewed: (i) first principles approaches, (ii) empirical model identification, and (iii) a hybrid approach that combines minimum metabolic network knowledge with an objective function to yield a predictive model.

In the development of the mathematical models described in this section, a number of "guiding principles" direct the organization of the elements of the biophysical networks. For purposes of nomenclature consistency and overall model organization, the constitutive elements of the biological hierarchy are reviewed briefly (expanding beyond the 3 simple levels described in the previous section). The following classification is adopted (Savageau, 2001):

- **Transcriptional Unit (TU):** lowest level element in hierarchy, input elements bind to receptor molecules and the output is an mRNA product.
- **Input signaling:** element which enables the transfer of an input signal to the TU, such as two component systems (regulator and receptor), combinations of extracellular and intracellular transport to regulator, as well as direct connection from another TU.
- **Mode:** category of response generation, either stimulation of response against a silent background, or repression of an otherwise active gene.
- **Logical Unit:** combined logic of multiple modes, referred to as *motifs* in the proceeding discussion.
- **Expression cascades:** also know as "central dogma", cascade of DNA to mRNA to protein product to metabolite product, with possibility of multi-stage, complex connectivity.
- **Connectivity:** overall architecture describing coordination of expression of related function (examples include operons and regulons).

To give an indication of complexity, the dimension of logical units for *E. coli* are on average 2-3 modulators per promotor site, and can be as high as 5 per site. On the output side (connectivity), TUs can have as many as 50 output connections (Savageau, 2001).

As emphasized in the introduction, an important point in systems biology is the *integrative* perspective, that is to say, the analysis of the system considered as a whole, and not the reductionist analysis of individual components. So while it is useful to categorize the elements and levels of a hierarchical regulatory scheme, it is more useful to analyze such schemes for behaviors that emerge from combinations of motifs. Some simple examples of canonical regulatory constructs that yield specific classes of behavior in gene networks include:

- **positive feedback:** multi-stability, oscillations, state-dependent response
- **integral feedback:** robust adaptation
- **negative feedback:** steady-state (homeostasis, adaptation)
- **time delay:** complex response, oscillations
- **protein oligomerization:** multi-stability, oscillations, resonant stimulus frequency response
- **stochastic fluctuations:** random response to stimuli, random outcomes, as well as stochastic focusing.

The interested reader is referred to, for example, (Smolen *et al.*, 2000) for additional insights.

2.2.1. Mechanistic Models Given detailed knowledge of a biological architecture, one can construct mathematical models to describe the behavior of interconnected motifs or TUs. A number of excellent review papers have been detailed in recent years (Smolen *et al.*, 2000; Hasty *et al.*, 2001*a*). In the majority of these studies, gene expression is described as a continuous time biochemical process, using combinations of algebraic and ordinary differential equations (ODEs) (Goldbeter, 1996; Smolen *et al.*, 2000; Cherry and Adler, 2000). In a similar manner, models at the signal transduction pathway level have been developed in a continuous time framework, yielding ODEs (Kholodenko *et al.*, 1999). At the TU level, a detailed mathematical treatment of transcriptional regulation is described in (Barkai and Leibler, 2000).

Mechanistic models for a number of specific biological systems have been reported, including basic operons and regulons in *E. coli* (trp, lac, pho), and bacteriophage systems (T7, λ) (see, for example, review in (Gilman and Arkin, 2002)). A detailed benchmark problem is described in (Zak *et al.*, 2003*a*) which consists of four biophysical motifs: cascade, mutual repression, autoactivation and sequestration, and agonist-induced receptor down-regulation. The complete network model consists of 118 reactions with 44 species and 97 parameters (rates of dimerization, degradation, transcription, translation, etc.). The latter model will be utilized in a subsequent discussion on identifiability.

2.2.2. Empirical Models Complementing the mechanistic methods for model development are the empirical or data-driven methods that include singular value decomposition analysis of microar-

ray data (Holter *et al.*, 2000; Alter *et al.*, 2000), self-organizing maps (Tamayo *et al.*, 1990), K-tuple means clustering or hierarchical clustering (Wen *et al.*, 1998; Spellman *et al.*, 1998; Eisen *et al.*, 1998), protein correlation and dynamic deviation factors (You and Yin, 2000), and robust statistics approaches (Zhao *et al.*, 2001; Thomas *et al.*, 2001). In contrast with the mechanistic approaches, most empirical approaches employ discrete-time gray box models (D'Haeseleer *et al.*, 1999; Weaver *et al.*, 1999; Wessels *et al.*, 2001; Hartemink *et al.*, 2002). A number of challenges are present, however, in treating experimental data for such problems: (i) the sampling rate is rarely uniform, and may be exponentially spaced by design, and (ii) data from multiple research groups is often combined (e.g., from WWW-posted data) to yield data records with inconsistent sampling, experimental bias, etc. From a systems engineering perspective, another critical point is the potentially divergent qualitative behavior between continuous time and discrete time models of corresponding order (Pearson, 1999). Recent work has shown the promise of continuous time formulations of empirical models using modulating function approaches (Zak *et al.*, 2003*b*).

There are a number of issues that must be considered in empirical modeling formulations. Many are related to notions of "identifiability" and design of suitable experiments. Unstructured approaches to model identification are completely ill-posed when faced with organisms such as the yeast cell with 6200 genes having four possible states (off, low, medium, high), leading to an overall expression state dimension in excess of $1E15$ (Lockhart and Winzler, 2000)! Clearly a number of *a priori* constraints and correlations must be exploited, as well as multiple sources of data.

A related challenge is the suitable design of rich experiments. In the case of yeast, repeated single gene knockouts of all genes leads to 6200×6200 profiles, without consideration for multiple knockouts. Mere extrapolation of current technology for microarrays will not solve these high dimensional data issues, however, systems engineering concepts of "rich" sequences can be exploited to improve the limited data. Perturbations can also be designed strategically. Typical knockouts involve so-called "direct effects" in which the expression level of various genes are altered in a network arrangement that involves direct connectivity to cis-regulatory elements of downstream genes (possible multiple cascades). An "indirect effect" can also be used (Davidson *et al.*, 2002) in which a mediating component (e.g., mRNA) is introduced to correct an intermediate element in the direct action cascade described previously.

2.2.3. Optimization-based Models A third category of modeling regulation can be considered to be a hybrid combination of empirical modeling and mechanistic modeling, and invokes principles of optimal control theory. The underlying assumption is that cells have been organized over evolutionary time scales to optimize their operations in a manner consistent with mathematical principles of optimality.

The cybernetic approach developed by Ramkrishna and co-workers (Kompala *et al.*, 1986; Varner and Ramkrishna, 1998) is founded on a simple principle; evolution has programmed or conditioned biological systems to optimally achieve physiological objectives. This straightforward concept can be translated into a set of optimal resource allocation problems that are solved at every time step in parallel with the model mass balances (basic metabolic network model). Thus, at every instant in time, gene expression and enzyme activity is rationalized as choice between sets of competing alternatives each with a relative cost and benefit for the organism. Mathematically, this can be translated into an instantaneous objective function. The researchers in this area have defined several postulates for specific pathway architectures, and the result is a computationally tractable (i.e., analytical) model structure. The potential shortcoming is a limited handling of more flexible objective functions that are commonly observed in biological systems (Bonarious *et al.*, 1997; Savinell and Palsson, 1992*a*; Savinell and Palsson, 1992*b*; Varma and Palsson, 1993*a*; Varma and Palsson, 1993*b*).

An alternative approach is the Flux Balance Analysis (FBA) (Watson, 1986), in which a suitable linear programming problem is posed and solved (Edwards *et al.*, 1999). The resulting model is not a dynamic model, and does not yield an analytical formulation, but the computational solution time is modest, and the approach has yielded success for a number of biological examples. Essential to the development of the model are the formulation of the system constraints, consisting of: (i) stoichiometric constraints that represent flux balances; (ii) thermodynamic constraints to restrict the directional flow through enzymatic reactions; and (iii) physicochemical capacity constraints to account for maximum flux through individual reactions. Recent extensions have addressed the problem of regulation by including additional time-dependent constraints in the formulation. The incorporation of transcriptional regulatory events in the FBA framework has extended the validity of the methodology for a number of complex dynamic system responses (Covert *et al.*, 2001).

In general, FBA has proven effective in applications where the steady state assumption is valid. However, there are many situations for which the steady state assumption is not valid, many of which are biophysically meaningful, such as the diauxic shift in *E. coli* . Dynamic extensions of the FBA algorithm have been proposed in (Mahadevan *et al.*, 2002) and will be described in a later section.

2.3 Analysis of Feedback Control Architectures

Control theory has found an enabling role in the analysis of the complex mathematical structures that result from the previously described modeling approaches. The language of control theory now dominates the quantitative characterization of biological regulation, as robustness, complexity, modularity, feedback, and fragility are invoked to describe these systems. Even classical control theoretic results such as the Bode sensitivity integral are being applied to describe the inherent tradeoffs in sensitivity across frequency (Csete and Doyle, 2002). Robustness has been introduced as both a biological system-specific attribute, as well as a measure of model validity (Ma and Iglesias, 2002). In the sections that follow, brief accounts are given of control theoretic analysis of biological regulatory structures, emphasizing where new insights into biological regulation have been uncovered.

2.3.1. Chemotaxis The signal transduction system that mediates chemotaxis exhibits a type of adaptation in which the response to a persistent stimulus is reset to the pre-stimulus value, thereby enabling an enhanced sensitivity. For a number of years, researchers speculated on a mechanistic explanation for this robust behavior, and two hypotheses had emerged: (i) precise fine tuning of several parameters to yield a consistent (robust) response under varied conditions, or (ii) inherent regulation that yielded this robust behavior. John Doyle and co-workers at Caltech were able to utilize the internal model principle to demonstrate that the regulatory system was exploiting integral feedback control to achieve the robust level of adaptation exhibited in chemotaxis, and more generally in systems with such behavior (Yi *et al.*, 2000). This understanding suggests that many seemingly complex biological networks may employ redundancy and other structural motifs or modules (enumerated in an earlier section) to achieve relatively simple overall system behavior (Lauffenburger, 2000).

2.3.2. Circadian Rhythm The gene network underlying circadian rhythm in the fly and in mammals has been the focus of detailed analysis in recent years (Goldbeter, 1996; Reppert, 2000; Winfree, 2001; Young and Kay, 2001; Goldbeter, 2002). The biological details are coming into sharper focus, as new experiments yield clues to the detailed (and somewhat overlapping) molecular circuitry of both the fly and mammals (Panda *et al.*, 2002). Building upon the evolving biological knowledge, there have been many postulated mathematical models (Leloup and Goldbeter, 1998; Tyson *et al.*, 1999; Scheper *et al.*, 1999; Lema *et al.*, 2000; Smolen *et al.*, 2001) that range in complexity from simple 2-state oscillators to more biophysically detailed transcriptional feedback schemes. As with adaptation in chemotaxis, *robustness* is the dominant characteristic often associated with the circadian rhythm regulatory loop (see, e.g., (Vilar *et al.*, 2002)), although formal systems-theoretic treatment of this behavior is a notable absence among the published reports. In a subsequent section, formal methods will be introduced for analyzing the robustness in the circadian rhythm gene network.

2.3.3. Stress Response As a final example, stress response is introduced to motivate both the ideas of reverse engineering as well as forward engineering. Cells are inherently robust to stochastic perturbations, and can readily recover from short-term exposure to environmental stressors (heat, pH, nutrient deprivation, etc.). The process of increasing specific protein activities to compensate for cellular insults has been termed stress response. One specific form of stress response that has been studied extensively is the cytoplasmic heat shock response in bacteria (see, e.g., (El-Samad *et al.*, 2002)). It has been established that bacteria respond to environmental perturbations such as an increase in temperature by rapidly inducing the synthesis of so-called heat shock proteins (hsps). These proteins play a critical role in protein processing, leading eventually to a restored balance in the organism. In *E. coli*, this balance is achieved through a feedback architecture involving so-called σ-factors, and the robust regulation of the transcription of the heat shock proteins. A recent mathematical analysis of the system (El-Samad *et al.*, 2002) has revealed the presence of nested inner and outer feedback loops, as well as feed-forward loops – suggesting that classical engineering control elements enable the robust regulatory behavior. Such understanding, or unraveling, is often termed *reverse engineering.*

In contrast, *forward engineering* involves the redesign of a regulatory system for a specific purpose or application. This can also be motivated in the context of stress response. One consequence of stress response is slower cell growth and decreased protein expression in an attempt to avoid

cell death. Such a response is particularly undesirable in the application of heterologous protein expression. This stressor has clear implications for biotechnology, as the high level expression of any protein is difficult and unpredictable. Our work to date suggests commonalities as well as key differences between this *endogenous* stress source in comparison with the more traditionally studies *exogenous* or environmental stressors. Modeling studies (Kauffmann *et al.*, 2000) are complemented with experimental studies employing *S. cerevisiae* to express a model single-chain antibody (scFv) protein (Kauffman *et al.*, 2002). Our ongoing work aims to understand cellular function in the context of this dynamic network, and focuses specifically on one regulatory network within the cell – the stress response to unfolded protein accumulation in the endoplasmic reticulum, or the unfolded protein response (UPR) (Travers *et al.*, 2000; Spear and Ng, 2001). In addition to reverse engineering the feedback architecture of the UPR for biological understanding, we aim to forward engineer cells that achieve higher levels of protein expression under the influence of a modified UPR.

3. SYSTEMS RESEARCH CHALLENGES IN SYSTEMS BIOLOGY

3.1 Robustness Analysis

Robustness is a recurring theme in the behavior of complex systems, whether they are man-made or natural (Morohashi *et al.*, 2001; Kitano, 2002). From an engineering perspective, one can interpret robustness as the ability to maintain some target level of behavior or performance in the presence of perturbations. In biological systems, these disturbances can be environmental (heat, pH, etc.) or intrinsic to the organism (changes in kinetic parameters). Recently, a framework has been established that elucidates the principle that the presence of robustness in a complex system requires the offsetting presence of fragility in the system (Csete and Doyle, 2002). While preliminary results are available for simple (low-dimensional, deterministic) systems, general tools for analyzing these tradeoffs are the subject of active research.

As mentioned previously, the gene network which underlies circadian rhythms is an ideal system to study robustness, owing to its remarkable performance in a highly uncertain environment. Of interest for control theoretic analyses, the dominant elements of the postulated architecture for *Drosophila* consist of nested negative autoregulatory feedback loops controlling the expression of timeless (*tim*) and period (*per*) interlocked with a positive feedback loop established via the *dClock* gene. Complex formation, regulated translocation and degradation of several of these gene products, which is additionally controlled (and delayed) by protein phosphorylation, add further levels of complexity to the system (Panda *et al.*, 2002).

A number of straightforward perturbation sensitivity studies have pointed to the robust character of this network under parameter uncertainty and stochastic variation owing to discrete molecular behavior (Leloup and Goldbeter, 1999; Barkai and Leibler, 2000; Smolen *et al.*, 2001; Gonze *et al.*, 2002). Our preliminary work addressed the variations in robustness properties among the very early, and relatively simple, models for this gene network. This work pointed to the presence of fragilities that were specific to different model classes (e.g., degradation fragility versus transcription fragility), as well as the convergence of molecular behavior at different rates to the continuum (i.e., deterministic) limit.

A more detailed treatment of this system has involved the use of the Fisher Information Matrix (FIM) to characterize points of relative sensitivity and robustness in the network (Stelling *et al.*, 2003). The FIM analysis provides a lower bound on parameter estimation accuracy, thus indicating robust elements (large variances) and fragile elements (tight variances). Corresponding to such a characterization are parametric sensitivities, which are high for fragile elements, and low for robust elements. Previous work had shown the utility of this approach for analyzing robustness in complex biophysical networks (Stelling and Gilles, 2001).

A model for *Drosophila* circadian rhythm was employed that consisted of 10 states, and 38 parameters to describe two branches of negative feedback regulation (*per*, *tim* loops). Two versions of the model were considered: the originally published deterministic model, and a stochastic model which accounts for the molecular interactions of small numbers of proteins. Gillespie's method was used to simulate the stochastic model (Gillespie, 1976). Parameters were varied to explore diverse areas in the operating space, and the rank-ordered sensitivity results are plotted in Figure 1. A striking characteristic of this plot is the uniform ordering of ranked sensitivities, despite the large variations explored in the parameter space. Two curves are depicted: the open circles show the results of the deterministic simulation, the lines connect the corresponding results for the formal stochastic simulation. The latter contains error bars to denote standard deviations in the ranges observed. The most sensitive or fragile parameters (lower rank order) correspond to global elements of the cellular machinery that are not specific to circadian rhythm, such as maximum rates of tran-

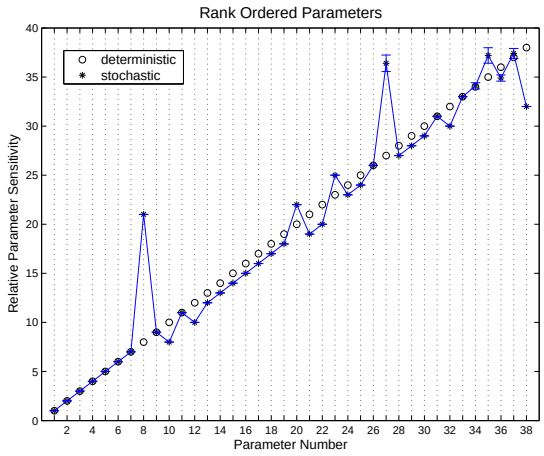

Fig. 1. Robustness of ordered sensitivity based on FIM calculations for all model parameters rank-ordered according to their accuracy.

scription. The robust parameters are those that correspond to the right side of the x-axis (high rank for sensitivity) and are associated primarily with the circadian rhythm specific attributes of the system such as the phosphorylation steps in the protein feedback paths. This suggests a design principle in which robust behavior in the circadian rhythm character is achieved at the expense of fragilities which are relegated to the central cellular machinery. A more detailed treatment of the analyses is described in (Stelling *et al.*, 2003).

These preliminary studies support the idea of robustness/fragility tradeoffs, and point to systems tools that can generate these insights. These tools are suitable for more complex problems, and with suitable modifications, can be applied to complex stochastic systems.

3.2 Identifiability of High Throughput Biological Data

It is anticipated that regulatory networks governing diverse cellular behavior will be discovered through the analysis of functional genomic data. Computational models may play an essential role in this task, as they can generate *in silico* data for validating analysis methods (Zak *et al.*, 2001; Smith *et al.*, 2002), and their properties may be analyzed to gain system–level insights. In this section, local identifiability analysis is applied to a computational model of a genetic regulatory network. The objective is to determine whether or not it is theoretically possible to uncover the network architecture using microarray (gene expression) data. It was observed that identifying the network architecture may only be possible when a rich microarray time course is coupled with information that specifies which transcription factors bind to which genes. The model used in the present study (Figure 2) is from a larger model that has been reported previously (Zak *et al.*, 2001). It is based on agonist–induced down–regulation of a steroid receptor (Brivanlou and Darnell, 2002). A soluble ligand (Q) diffuses through the plasma membrane and into the nucleus where it binds to and activates a transcription factor (the steroid receptor, E), which then causes changes in the expression of target genes (F). In this model, these changes in gene expression ultimately lead to down–regulation of the steroid receptor.

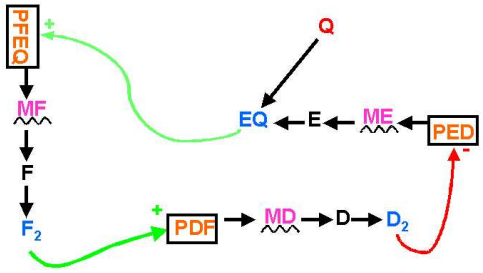

Fig. 2. Model of genetic regulatory network

For the present analysis, this system is represented by a set of ordinary differential equations (ODEs). Transcriptional regulation is modeled in the manner of (Barkai and Leibler, 2000). Equation (1) shows the equations pertaining to D: promoter (PDF) binding by a transcription factor (F_2), transcription, translation of the transcript (MD), dimerization of the protein (D), and further promoter binding. The other components are described similarly.

$$\begin{aligned}
[\dot{PDF}] &= -k_{PDF}[PDF][F_2] + k_{UPDF}[F_2PDF] \\
[\dot{F_2PDF}] &= k_{PDF}[PDF][F_2] - k_{UPDF}[F_2PDF] \\
[\dot{MD}] &= k_{RPDF}[PDF] + k_{RFPDF}[F_2PDF] - k_{dMD}[MD] \\
[\dot{D}] &= k_{TD}[MD] - 2k_{D2}[D]^2 + 2k_{UD2}[D_2] - k_{dD}[D] \\
[\dot{D_2}] &= k_{D2}[D]^2 - k_{UD2}[D_2] - k_{dD2}[D_2] \\
&\quad -k_{PED}[PED][D_2] + k_{UPED}[D_2PED]
\end{aligned} \tag{1}$$

The overall model consists of 13 states and 31 parameters. Parameter values were taken from similar biological systems in the literature. Sample time courses of the transcripts (MD, ME, and MF) in response to ligand inputs are shown in Figure 3.

Initially, the problem of identifying the architecture of the system in Figure 2 from expression profiles of MD, ME, and MF for cases with and without *localization information* is considered. Localization information specifies which transcription factors bind to which promoters (for example (Simon *et al.*, 2001)). When there is localization information available, the identification of the system architecture involves identifying the

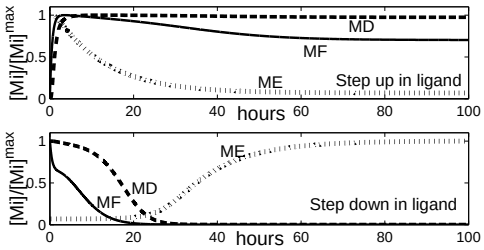

Fig. 3. Expression profiles in response to a step up (*top*) and step down (*bottom*) in ligand (Q) concentration.

transcription rate constants for the bound and unbound promoters for each gene (k_{RPDF} and k_{RFPDF} for gene D, for example), six parameters in total. In the case without localization information, identifying the architecture involves identifying the binding constants for each transcription factor for each gene as well as the transcription rate constants, giving a total of 27 parameters ($3\times3 + 6\times3$), of which only six are non–zero. The difference is illustrated in Figure 4, where each $\mathbf{X}$ corresponds to 3 parameters (1 promoter binding and 2 transcription rate constants) that must be identified. A model parameter is *locally identifiable* if a set of ideal measurements determines a finite set of values for that parameter (Jacquez and Perry, 1990). Local identifiability is weaker, but more computationally feasible, than global identifiability, which requires unique determination of parameter values. If a parameter is not locally or globally identifiable, its value can only be determined as a combination of other parameters. For this reason, identifiability analysis is a key component of parameter estimation. Following (Jacquez and Perry, 1990), local identifiability analysis involves first determining the $N_s \times N_p$ sensitivity matrix, $S(t)$ for the ODE system of N_s states (x) and N_p parameters (p). It is rare to measure all of the dynamic states (x) in an experiment, and thus the sensitivities of the measurements, $S'(t)$, may be given by $S'(t) = CS(t)$, where the measurements (y) are related to the states by $y(t) = Cx(t)$. Values of S' at several different times may be stacked to give G:

Fig. 4. Possible transcription factor–gene interactions in model, demonstrating the constraining effect of localization. (a) Localization is not known. (b) Localization is known.

$$G \equiv [S'(t_1), S'(t_2), ...]^T \qquad (2)$$

The system is locally identifiable if $det(G^TG) \neq 0$. When $det(G^TG) = 0$, putting it into reduced row echelon form will reveal rows that contain single nonzero elements. The column indices of these elements indicate locally identifiable parameters. Rows containing multiple nonzero elements indicate parameters that are not locally identifiable (Jacquez and Perry, 1990). Note that when this analysis is applied to nonlinear systems, absence of local linear identifiability does not guarantee absence of local identifiability for the nonlinear system, while the reverse is true for global identifiability (Jacquez and Perry, 1990).

The local identifiability analysis is performed for cases where microarray data is used alone or with localization information for three experimental designs of ligand step up, step down, and combined steps. Results are shown in Figure 5. If microarray data is used alone, the parameters that determine the network architecture are not locally identifiable. Inclusion of localization information allows the network architecture parameters for gene E to become locally identifiable for either of the steps (step down not shown). Only when both steps are used together with localization information do all of the network architecture parameters become identifiable. For this case, there are also many more types of parameters that become locally identifiable.

(a) *STEP UP ALONE*

PARAMETER TYPE	NUMBER OF IDENTIFIABLE PARAMS. *microarray alone*	*microarray & localization*
promoter binding	4/9	1/3
transcription	0/18	4/6
transcript degradation	0/3	3/3
protein degradation	0/3	1/3
dimer degradation	0/3	1/3

(b) *STEP UP & STEP DOWN*

PARAMETER TYPE	NUMBER OF IDENTIFIABLE PARAMS. *microarray alone*	*microarray & localization*
promoter binding	4/9	3/3
transcription	0/18	6/6
promoter unbinding	0/9	2/3
transcript degradation	0/3	3/3
protein degradation	0/3	3/3
dimer degradation	0/3	3/3
dimerization	0/3	2/3
translation	0/3	2/3
ligand degradation	1/1	1/1

Fig. 5. *Localization and rich data are necessary to determine network architecture.* Locally identifiable parameters when microarray data is used alone or with localization information. (a) Step up in ligand alone. (b) Step up and step down used together. Emphasized parameters are those necessary for determining network architecture.

Even with localization information, the network architecture is identifiable only when both steps were used together. These results demonstrate the importance of *rich* experimental data. The microarray data that may be acquired in the near future, however, may not be rich enough for determination of network architecture. Even with rich

data, in the absence of localization information, the parameters that determine network architecture are not identifiable. This motivates combining microarray data with other information that constrains possible interactions, and is consistent with previous studies that demonstrated that inclusion of localization information improved determination of regulatory network architectures (Zak *et al.*, 2001; Hartemink *et al.*, 2002).

3.3 Optimality Approaches to Modeling Dynamic Regulation

As described in the modeling section, an interesting intersection of systems theory and regulation modeling is in optimization-based approaches such as FBA or cybernetic models. The cybernetic modeling framework is predicated on a specific objective formulation, namely an instantaneous objective, point-wise in time. FBA, on the other hand, accommodates a more general cost formulation, but does not yield an analytical model and yields an inherently steady-state model. In this section, the details of a dynamic extension to the FBA are highlighted. Specifically, two formulations are considered: a horizon-based objective (dynamic optimization approach), and a static optimization approach.

Dynamic Optimization based dFBA Approach (DOA) Consider a metabolic network with m metabolites and n fluxes. The set of conservation of mass equations, for each metabolite, results in a set of ordinary differential equations;

$$\begin{aligned} \frac{d\boldsymbol{z}}{dt} &= A\boldsymbol{v}X \\ \frac{dX}{dt} &= \mu X \\ \mu &= \Sigma w_i v_i \end{aligned} \tag{3}$$

where $\boldsymbol{z}$ is the vector of metabolite concentrations, X is the biomass concentration, $\boldsymbol{v}$ is the vector of metabolic fluxes per gram (DW) of the biomass, A is the stoichiometric matrix of the metabolic network, μ is the growth rate obtained as a weighted sum of the reactions that synthesize the growth precursors, and w_i are the amounts of the various growth precursors required per gram (DW) of biomass.

Along with the system of dynamic equations, several additional constraints must be imposed for a realistic prediction of the metabolite concentrations and the metabolic fluxes. These include: non-negative metabolite and flux levels, limits on the rate of change of fluxes, and any additional nonlinear constraints on the transport fluxes. A general dynamic optimization problem can be formulated as shown in Equation 4:

$$\begin{aligned} \underset{\boldsymbol{z}(t),\boldsymbol{v}(t),X(t)}{\text{Max.}} \quad & \hat{w}_{end}\Phi(\boldsymbol{z},\boldsymbol{v},X)|_{t=t_f} + \\ & \hat{w}_{ins}\sum_{j=0}^{M}\int_{t_0}^{t_f} L(\boldsymbol{z},\boldsymbol{v},X(t))\delta(t-t_j)dt \\ \text{s.t.} \quad & \frac{d\boldsymbol{z}}{dt} = A\boldsymbol{v}X \\ & \frac{dX}{dt} = \mu X \\ & \mu = \Sigma w_i v_i \\ & t_j = t_0 + j\frac{t_f - t_0}{M} \quad j = 0 \ldots M \\ & c(\boldsymbol{v},\boldsymbol{z}) \leq 0 \quad |\dot{\boldsymbol{v}}| \leq \dot{\boldsymbol{v}}_{max} \\ & \boldsymbol{z} \geq 0 \quad X \geq 0 \\ & \boldsymbol{z}(t_0) = \boldsymbol{z_0} \quad X(t_0) = X_0 \end{aligned} \tag{4}$$

where $\boldsymbol{z_0}$ and X_0 are the initial conditions for the metabolite concentration and the biomass concentration respectively, $c(\boldsymbol{v},\boldsymbol{z})$ is a vector function representing nonlinear constraints that could arise due to consideration of kinetic expressions for fluxes, t_0 and t_f are the initial and the final time, Φ is the terminal objective function that depends on the end-point concentration, L is the instantaneous objective function, δ is the Dirac-delta function, t_j is the time instant at which L is considered, $\hat{w}_{ins}$ and $\hat{w}_{end}$ are the weights associated with the instantaneous and the terminal objective function respectively and $\boldsymbol{v}(t)$ is the time profile of the metabolic fluxes. If the nonlinear constraint is absent, the problem reduces to an optimization involving a bilinear system.

The dynamic optimization problem can be solved by parameterizing the dynamic equations through the use of orthogonal collocation on finite elements (Cuthrell and Biegler, 1987). The time period (t_0-t_f) is divided into a finite number of intervals (finite elements). The fluxes, the metabolite levels and the biomass concentration are parameterized at the roots of an orthogonal polynomial within each finite element. The details of the parameterization for a specific example are presented in (Mahadevan *et al.*, 2002).

Static Optimization based dFBA Approach (SOA) In SOA, the time period is divided into N intervals. In the absence of the nonlinear constraints involving the fluxes, the optimization problem is reduced to a LP problem. The LP (equation 5) is solved at the beginning of each interval to obtain the fluxes at that time instant:

$$
\begin{aligned}
\underset{\boldsymbol{v}(t)}{\text{Max.}} \quad & \Sigma w_i v_i(t) \\
\text{s.t.} \quad & \boldsymbol{z}(t+\Delta T) \geq 0 \quad \boldsymbol{v}(t) \geq 0 \\
& \hat{c}(\boldsymbol{z}(t))\boldsymbol{v}(t) \leq 0 \quad \forall t \in [t_0, t_f] \\
& |\boldsymbol{v}(t) - v(t-\Delta T)| \leq \dot{\boldsymbol{v}}_{max} \Delta T \\
& \boldsymbol{z}(t+\Delta T) = \boldsymbol{z}(t) + A\boldsymbol{v}\Delta T \\
& X(t+\Delta T) = X(t) + \mu X(t)\Delta T
\end{aligned} \tag{5}
$$

where ΔT is the length of the time interval chosen.

The dynamic equations are integrated assuming that the fluxes are constant over the interval. The optimization problem is then formulated at the next time instant and solved. This procedure is repeated from t_0 to t_f. For the class of systems involving only bilinear terms with fluxes and the biomass concentration, it is possible to directly solve the dynamic equations and thereby eliminate the numerical integration.

A detailed description of the dFBA approach and its application to metabolic modeling can be found in (Mahadevan *et al.*, 2002). The pertinent observation is the traditional FBA modeling approach invokes a steady state assumption that allows one to solve for the metabolic flux distribution, but eliminates the ability to track metabolic concentrations or study metabolic transients. The dFBA extension allows the tracking of metabolite concentrations and dynamic transients. In effect, regulation is captured by virtue of the optimization objective solved under the constraint of dynamic metabolic fluxes.

4. SUMMARY

Biological regulation has been introduced and analyzed from the perspective of systems engineering. Mathematical modeling approaches, both empirical and fundamental, have yielded descriptions of many complex systems, and control-theoretic tools have been employed to provide hypotheses for biological behavior, such as system robustness. Open challenges were described in the areas of robustness analysis, design of experiments, and mathematical modeling frameworks. Future advances will require complementary approaches from experimental disciplines as well as theoretical and computational disciplines in order to achieve the aims of systems biology.

5. ACKNOWLEDGMENTS

A number of ideas in this paper were formed over the course of a sabbatical visit at Stuttgart University. The author would like to acknowledge his host (Prof. Frank Allgöwer), funding sponsor (Alexander von Humboldt Foundation), and stimulating collaboration with Jörg Stelling at the Max Planck Institute in Magdeburg. The dFBA work is a collaboration with Prof. Jeremy Edwards (U. Delaware), and the UPR work is collaborative with Prof. Anne Robinson (U. Delaware). The author would also like to acknowledge the work of his present and former students that contributed to this body of work, notably Nada Bagheri, Radhakrishnan Mahadevan, and Dan Zak.

REFERENCES

Alter, O., P.O. Brown and B. Botstein (2000). Singular value decomposition for genome wide expression data processing and modeling. *Proc. Natl. Acad. Sci. USA* **97**, 10101–10116.

Barkai, N. and S. Leibler (2000). Circadian clocks limited by noise. *Nature* **403**, 267–8.

Bonarious, H.P.J., G. Schmid and J. Tramper (1997). Flux analysis of underdetermined metabolic networks: The quest for the missing constraints. *Trends Biotech.* **15**, 308–314.

Brivanlou, A. H. and J. E. Darnell (2002). Signal transduction and the control of gene expression. *Science* **295**, 813–818.

Cherry, J L. and F R. Adler (2000). How to make a biological switch. *J Theor Biol* **203**, 117–133.

Covert, M.W., C.H. Schilling and B. Palsson (2001). Regulation of gene expression in flux balance models of metabolism. *J. Theor. Biol.* **213**, 73–88.

Csete, M.E. and J.C. Doyle (2002). Reverse engineering of biological complexity. *Science* **295**, 1664–1669.

Cuthrell, J. E. and L.T. Biegler (1987). On the Optimization of Differential Algebraic Process Systems. *AIChE J.* **33**(8), 1257–12700.

Davidson, E.H., J.P. Rast, P. Oliveri, A. Ransick, C. Calestani, C-H. Yuh, T. Minokawa, G. Amore, V. Hinman, C. Arenas-Mena, O. Otim, C.T. Brown, C.B. Livi, P.Y. Lee, R. Revilla, A.G. Rust, Z.J. Pan, M.J. Schilstra, P.J.C. Clarke, M.I. Arnone, L. Rowen, R.A. Cameron, D.R. McClay, L. Hood and H. Bolouri (2002). A genomic regulatory network for development. *Science* **295**, 1669–1678.

D'Haeseleer, P, X Wen, S Fuhrman and R Somogyi (1999). Linear modeling of *mRNA* expression levels during CNS development and injury. In: *Pacific Symposium on Biocomputing 4*. pp. 41–52.

Edwards, J.S., R. Ramakrishna, C. H. Schilling and B. O. Palsson (1999). Metabolic Flux Balance Analysis. In: *Metabolic Engineering* (S. Y. Lee and E. T. Papoutsakis, Eds.). pp. 13–57. Marcel Deker.

Eisen, M.B., P.T. Spellman, P.O. Brown and D. Botstein (1998). Cluster analysis and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95**, 14863–14868.

El-Samad, H., M. Khammash, H. Kurata and J. Doyle (2002). Robustness analysis of the heat shock response in *E. coli*. In: *Proc. Amer. Control Conf.*. p. 17421747.

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

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

Gilman, A. and A. Arkin (2002). Genetic "code": Representations and dynamical models of genetic components and networks. *Annu. Rev. Genomics Hum. Genet.* **3**, 341–369.

Goldbeter, A. (1996). *Biochemical oscillations and cellular rhythms: the molecular bases of periodic and chaotic behaviour*. 2nd ed.. Cambridge Univ. Press.

Goldbeter, A. (2002). Computational approaches to cellular rhythms. *Nature* **420**, 238–245.

Gonze, D., J. Halloy and A. Goldbeter (2002). Robustness of circadian rhythms with respect to molecular noise. *Proc. Natl. Acad. Sci. USA* **99**(2), 673–78.

Hartemink, A. J., D. K. Gifford, T. S. Jaakola and R. A. Young (2002). Combining location and expression data for principled discovery of genetic regulatory network models. In *Proc. Pac. Symp. Biocomput.* **7**, 437–49.

Hasty, J., D. McMillen, F. Isaacs and J.J. Collins (2001*a*). Computational studies of gene regulatory networks: In numero molecular biology. *Nature Rev. Gen.* **2**, 268–279.

Hasty, J., F. Isaacs, M. Dolnik, D. McMillen and J.J. Collins (2001*b*). Designer gene networks: towards fundamental cellular control. *Chaos* **11**(1), 207–220.

Holter, N.S., M. Mitra, A. Maritan, M. Cieplak, J.R. Banavar and N.V. Fedoroff (2000). Fundamental patterns underlying gene expression profiles: Simplicity from complexity. *Proc. Natl. Acad. Sci. USA* **97**, 8409–8414.

Jacquez, J. A. and T. Perry (1990). Parameter estimation: local identifiability of parameters. *Am. J. Physiol.* **258**(4), E727–E736.

Judd, E.M., M.T. Laub and H.H. McAdams (2000). Toggles and oscillators: new genetic circuit designs. *BioEssays* **22**, 507–509.

Kauffman, K.J., E.M. Pridgen, F.J. Doyle III, P.S. Dhurjati and A.S. Robinson (2002). Decreased protein expression and intermittent recoveries in BiP levels result from cellular stress during heterologous protein expression in *Saccharomyces cerivisiae*. *Biotech. Progress* **18**, 942–950.

Kauffmann, K., P. Dhurjati, A.S. Robinson and F.J. Doyle III (2000). Framework for modeling information flow in biological processes: Application to the unfolded protein response. In: *Proc. IFAC Conf. Computer Appl. Biotechnology*. pp. 137–142.

Kholodenko, B N., O V. Demin, G Moehren and J B. Hoek (1999). Quantification of short term signaling by the epidermal growth factor receptor. *J. Biol. Chem.* **274**(42), 30169–30181.

Kitano, H. (2002). Systems biology: a brief overview. *Science* **295**, 1662–64.

Kompala, D.S., D. Ramkrishna, N.B. Jansen and G.T. Tsao (1986). Investigation of bacterial growth on mixed substrates. Experimental evaluation of cybernetic models. *Biotech. Bioeng.* **28**, 1044–1056.

Lauffenburger, D.A. (2000). Cell signaling pathways as control modules: Complexity for simplicity?. *Proc. Natl. Acad. Sci. USA* **97**, 5031–5033.

Lee, T.I., N.J. Rinaldi, F. RObert, D.T. Odom, Z. barJospeh, G.K. Gerber, N.M. Hannett, C.T. Harbison, C.M. THompson, I. Simon, J. Zeitlinger, E.G. Jennings, H.L. Murray, D.B. Gordon, B. Ren, J.J. Wyrick, J.B. Tagne, T.L. Volkert, E. Fraenkel, D.K. Gifford and R.A. Young (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* **298**, 799–804.

Leloup, J.-C. and A. Goldbeter (1998). A model for circadian rhythms in *Drosophila* incorporting the formation of a complex between the PER and TIM proteins. *J. Biol. Rhy.* **13**, 70–87.

Leloup, J.-C. and A. Goldbeter (1999). Chaos and birhythmicity in a model for circadian oscillations of the PER and TIM proteins in *Drosophila*. *J. Theor. Biol.* **198**, 445–59.

Lema, M. A., D. A. Golombek and J. Echave (2000). Delay model of the circadian pacemaker. *J. Theor. Biol* **204**, 565–573.

Lockhart, D.J. and E.A. Winzler (2000). Genomics, gene expression and DNA arrays. *Nature* **405**, 827–836.

Ma, L. and P.A. Iglesias (2002). Quantifying robustness of biochemical network models. *BMC Bioinformatics* **3**, 38–50.

Mahadevan, K., J. Edwards and F.J. Doyle III (2002). Dynamic flux balance analysis of diauxic growth in *Escherichia coli*. *Biophys. J.* **83**, 1331–1340.

Morohashi, M., A.E. Winn, M.T. Borisuk, H. Bolouri, J. Doyle and H. Kitano (2001). Robustness as a measure of plausibility in models of biochemical networks. *J. Theor. Biol.* **216**, 19–30.

Panda, S., J.B. Hogenesch and S.A. Kay (2002). Circadian rhythms from flies to human. *Nature* **417**, 329–335.

Pearson, R. K. (1999). *Discrete-Time Dynamic Models*. Oxford University Press.

Reppert, S.M. (2000). Comparing clockworks: Mouse versus fly. *J. Biol. Rhy.* **15**(5), 357–364.

Savageau, M.A. (2001). Design principles for elementary gene circuits: Elements, methods, and examples. *Chaos* **11**(1), 142–159.

Savinell, J.M. and B.O. Palsson (1992*a*). Network analysis of intermediary metabolism using linear optimization. Development of mathematical formalism. *J. Theor. Biol.* **154**, 421–454.

Savinell, J.M. and B.O. Palsson (1992*b*). Network analysis of intermediary metabolism using linear optimization. Interpretation of hybridoma cell metabolism. *J. Theor. Biol.* **154**, 455–473.

Scheper, T., D. Klinkenberg, C. Pennartz and J. van Pelt (1999). A mathematical model for the intracellular rhythm generator. *J. Neurosci.* **19**, 40–47.

Shen-Orr, S.S., R. Milo, S. Mangan and U. Alon (2002). Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nature Genetics* **31**, 64–68.

Simon, I., J. Barnett, N. Hannett, C. T. Harbison, N. J. Rinaldi, T. L. Volkert, J. J. Wyrick, J. Zeitlinger, D. K. Gifford, T S. Jaakkola and R A. Young (2001). Serial regulation of transcriptional regulators in the yeast cell cycle. *Cell* **106**, 697–708.

Smith, V A., E D. Jarvis and A J. Hartemink (2002). Evaluating functional network inference using simulations of complex biological systems. *Bioinformatics* **18 Suppl. 1**, S216–S224.

Smolen, P., D.A. Baxter and J.H. Byrne (2000). Mathematical modeling of gene networks. *Neuron* **26**, 567–580.

Smolen, P., D.A. Baxter and J.H. Byrne (2001). Modeling circadian oscillations with interlocking positive and negative feedback loops. *J. Neurosc.* **21**, 6644–56.

Spear, E. and D.T.W. Ng (2001). The unfolded protein response: No longer just a special teams player. *Traffic* **2**, 515–523.

Spellman, P.T., G. Sherlock, M.Q. Zhang, V.R. Iyer, K. Anders, M.B. Eisen, P.O. Brown, D. Botstein and B. Futcher (1998). Comprehensive identification of cell cycle-regulated genes of the yeast *Saccharomyces cerevisiae* by microarray hybridization. *Molec. Biol. Cell* **9**, 3273–3297.

Stelling, J. and E.D. Gilles (2001). Robustness vs. identifiability of regulatory modules?. In: *Proc. 2nd Intl. Conf. on Systems Biology*. pp. 181–190.

Stelling, J., E.D. Gilles, D. Zak and F.J. Doyle III (2003). Robust finetuning of the drosophila circadian clock. *Nature*. submitted.

Tamayo, P., D. Slonium, J. Mesirov, Q. Zhu, S. Kitareewan, E. Dmitrovsky, E. Lander and T.R. Golub

(1990). Interpreting patterns of gene expression with self-organizing maps: Methods and application to hematopoietic differentiation. *Proc. Natl. Acad. Sci. USA* **44**, 129–131.

Thomas, J.G., J.M. Olson, S.J. Tapscott and L.P. Zhao (2001). An efficient and robust statistical modeling approach to discover differentially expressed genes using genomic expression profiles. *Genome Research* **11**, 1227–1236.

Travers, K.J., C.K. Patil, L. Wodicka, D.J. Lockhart, J.S. Weissmann and P. Walter (2000). Functional and genomic analyses reveal an essential coordination between the unfolded protein response and er-associated degradation. *Cell* **101**, 249–258.

Tyson, J. J., C. I. Hong, C. D. Thron and B. Novak (1999). A simple model of circadian rhythms based on dimerization and proteolysis of per and tim. *Biophys. J.* **77**, 2411–2417.

Varma, A. and B.O. Palsson (1993*a*). Metabolic capabilities of *Escherichia coli*: I. Synthesis of biosynthetic precursors and cofactors. *J. Theor. Biol.* **165**, 477–502.

Varma, A. and B.O. Palsson (1993*b*). Metabolic capabilities of *Escherichia coli*: Ii. Optimal growth patterns. *J. Theor. Biol.* **165**, 503–522.

Varner, J. and D. Ramkrishna (1998). Application of cybernetic models to metabolic engineering: Investigation of storage pathways. *Biotech. Bioeng.* **58**, 282–291.

Vilar, J M., H Y. Kueh, N. Barkai and S. Leibler (2002). Mechanisms of noise-resistance in genetic oscillators.. *Proc. Natl. Acad. Sci. USA* **30**(99), 5988–92.

Watson, M.R. (1986). A discrete model of bacterial metabolism. *Comp. Appl. Biosciences* **2**, 23–27.

Weaver, D C., C T. Workman and G D. Stormo (1999). Modeling regulatory networks with weight matrices. In: *Pacific Symposium on Biocomputing 4*. pp. 102–111.

Wen, X., S. Fuhrman, G.S. Michaels, D.B. Carr, S. Smith, J.L. Barker and R. Somogyi (1998). Large-scale temporal gene expression mapping of central nervous system development. *Proc. Natl. Acad. Sci. USA* **95**, 334–339.

Wessels, L F.A., E P. Van Someren and M J.T. Reinders (2001). A comparison of genetic network models. In: *Pacific Symposium on Biocomputing 6*. pp. 508–519.

Winfree, A.T. (2001). *The Geometry of Biological Time*. 2nd ed.. Springer. New York.

Yi, T. M., Y. Huang, M. I. Simon and J. Doyle (2000). Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proc. Natl. Acad. Sci. USA* **97**, 4649–4653.

You, L. and J. Yin (2000). Patterns of regulation from mrna and protein time series. *Metabolic Eng.* **2**, 210–217.

Young, M. W. and S. A. Kay (2001). Time zones: a comparative genetics of circadian clocks. *Nat. Rev. Genet.* **2**, 702–715.

Zak, D E., F J. Doyle III, G E. Gonye and J S. Schwaber (2001). Simulation studies for the identification of genetic networks from cDNA array and regulatory activity data. In: *Proc. 2nd Intl. Conf. Systems Biology*. pp. 231–238.

Zak, D., G. Gonye, J.S. Schwaber and F.J. Doyle III (2003*a*). Importance of input perturbations and stochastic gene expression in the reverse engineering of genetic regulatory networks: Insights from an identifiability analysis of an in silico network. *Genome Research*. in press.

Zak, D., R.K. Pearson, R. Vadigepalli, G. Gonye, J.S. Schwaber and F.J. Doyle III (2003*b*). Continuous-time identification of gene expression models. *Omics*. in press.

Zhao, L.P., R. Prentice and L. Breeden (2001). Statistical modeling of large microarray data sets to identify stimulus-response profiles. *Proc. Nat. Acad. Sci. USA* **98**(10), 5631–5636.