Proceedings of the 17th World Congress
The International Federation of Automatic Control
Seoul, Korea, July 6-11, 2008

Robust Control in Biology: From Genes to Cells to Systems

Francis J. Doyle III

University of California, Santa Barbara, CA 93111, USA
(Tel: 805-893-8133; e-mail: doyle@engr.ucsb.edu)

Abstract: Challenges in the field of systems biology are outlined from the perspective of control and dynamical systems. These exquisite biophysical networks have enviable properties with regard to robustness to disturbances and uncertainty, as well as noise tolerance. Several examples are used to motivate the ideas, including neurons controlling circadian rhythms, programmed cell death (apoptosis), and signaling pathways for glucose metabolism.

1. INTRODUCTION

Natural control systems are paragons of optimality. Over millennia, these architectures have been honed to achieve robust regulation of a myriad of processes at the levels of genes, proteins, cells, and entire systems. One of the more challenging opportunities for "systems" research is unraveling the multi-scale, hierarchical control that achieves robust performance in the face of stochastic perturbations. These perturbations arise from both intrinsic sources (e.g., inherent variability in the transcription machinery), and extrinsic sources (e.g., environmental fluctuations). Robustness in key performance variables to particular perturbations has been shown to be achieved at the expense of strong sensitivity to other perturbations.

In this paper, several biological examples will be used to highlight robustly regulated behavior, including: circadian timekeeping in neuronal cells; the pathways underlying insulin resistance in diabetes; and programmed cell death (apoptosis). A key insight from these examples is that control at the cellular network level guides many properties in a manner that is distinct from control at the intracellular level, or even control at the organism level.

A variety of tools from systems theory are employed in this research, including the structured singular value, sensitivity measures (with extensions to limit cycle behavior and stochastic systems), and discrete stochastic simulations. Those tools complement the high throughput biological assays that are used to interrogate the natural control circuits.

2. SYSTEMS BIOLOGY

Advances in molecular biology over the past 2 decades have shed led on the relationships between 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 (Hasty *et al.,* 2001). Model development involves translating identified biological networks into coupled dynamical equations that are amenable to numerical simulation and analysis. These equations detail the complex biophysical processes that create interactions between the "nodes" in the network as well as with the external environment of the cell or the organism. These interactions involve hierarchical feedback loops that led to the robust system response to disturbances, including intrinsic noise, as well as environmental stressors.

The discipline of Systems Biology has emerged in response to these challenges (Ideker *et al.*, 2001; Kitano 2001), and combines approaches and methods from systems engineering, computational biology, statistics, genomics, molecular biology, biophysics, and other fields (Klipp *et al.*, 2005; Palsson 2006; Szallasi *et al.*, 2006). The recurring themes include: (i) integrative viewpoints towards unraveling complex dynamical systems, and (ii) tight iterations between experiments, modeling, and hypothesis generation.

2.1 Early Successes

Although the field of Systems Biology is relatively young, one can already point to early successes in a number of cases. The work of Adam Arkin on λ-phage was one of the first detailed analyses of a stochastic gene switch, and showed convincingly that formal stochastic treatment was required to understand the cell fate switch between lysis and lysogeny (Arkin *et al.,* 1998). The analysis of perfect adaptation in chemotaxis is another example where multiple groups adopted a "systems" perspective, and key insights have been generated (Barkai and Leibler, 1997; Rao *et al.,* 2004; Yi *et al.,* 2000). Notably, the mechanism for perfect adaptation has been elucidated and interpreted in classical control engineering terms: integral feedback (Yi *et al.,* 2000). The approach of model reduction and systematic analysis (including requisite modeling assumptions to yield perfect adaptation) is a perfect example of an effective "systems" strategy. This problem continues to generate new insights, as recent work has shown that disparate organisms have both

overlapping and distinctive architectures for chemotaxis. Another nice example that has received considerable attention is the gene network underlying circadian rhythm. Models have been proposed (e.g., Goldbeter 1996), and formal robustness analysis tools have generated insights on biological design principles. A more detailed case study that might be characterized as a "success story" has emerged from the work of Timmer and Klingmüller on the JAK-STAT pathway (Muller *et al.,* 2004). They have shown that modeling-experiment iterations can yield new hypothesis – particularly regarding unobservable components that can be simulated (but never measured). One implication, for the JAK-STAT pathway, involves pharmacological intervention. Current practice focuses on the phosphorylation element of the pathway, but the model shows that a more effective strategy involves the blocking of nuclear export.

2.2 Control and Systems Biology

In particular, the field of control engineering has had a pervasive influence on the discipline of systems biology. The chemotaxis work mentioned previously (Yi *et al.,* 2000) represented a collaboration between control engineers (J. Doyle) and biologists (Simon). The work of this author and his collaborators has also been at the interface of control and biology: notably the robustness analysis of cellular function (Stelling *et al.,* 2004b) and the unraveling of design principles in circadian rhythm (Stelling *et al.,* 2004a). More theoretical approaches have been taken by control mathematicians such as Sontag, with major advances in understanding signal transduction (Sontag 2002), and the oscillations underlying a positive feedback gene switch (Angeli *et al.,* 2004). There are many other contributions from the control community, and space precludes their complete enumeration here.

3. BIOLOGICAL NETWORKS

Biophysical networks are remarkably diverse, cover a wide spectrum of scales, and are inevitably characterized by a range of complex behaviors. These networks have attracted a great deal of attention at the level of gene regulation, where dozens of input connections may characterize the regulatory domain of a single gene in a eukaryote, as well as the protein level where literally thousands of interactions have been mapped in so-called protein interactome diagrams that illustrate the potential coupling of pairs of proteins (Barabasi, 2004; Malcom *et al*., 2003). However these networks also exist at higher levels, including the coupling of individual cells via signaling molecules, the coupling of organs via endocrine signaling, and ultimately the coupling of organisms in eco-systems. To elucidate the mechanisms employed by these networks, biological experimentation and intuition are by themselves insufficient. As noted earlier, the field of systems biology has laid claim to this class of problems, and engineers, biologists, physicists, chemists, mathematicians, and many others have united to embrace these problems with interdisciplinary approaches (Kitano, 2002). In this field, investigators characterize dynamics via mathematical models and apply systems theory with the goal of guiding further experimentation to better understand the biological network that gives rise to robust performance (Kitano, 2002).

3.1 Circadian Rhythm Network

An ideal example of such networked biological complexity is the circadian clock, which coordinates daily physiological behaviors of most organisms. The mammalian circadian master clock resides in the suprachiasmatic nucleus (SCN), located in the hypothalamus (Reppert *et al*., 2002). It is a network of multiple autonomous noisy oscillators, which communicate via neuropeptides to synchronize and form a coherent oscillator (Herzog *et al*., 2004; Liu *et al*., 2007). At the core of the clock is a gene regulatory network in which approximately 6 key genes are regulated through an elegant array of time-delayed negative feedback circuits (see Figure 1). The activity states of the proteins in this network are modulated (activated/inactivated) through a series of chemical reactions including phosphorylation and dimerization. These networks exist at the subcellular level. Above this layer is the signaling that leads to a synchromized response from the population of thousands of clock neurons in the SCN. Ultimately, this coherent oscillator then coordinates the timing of daily behaviors, such as the sleep/wake cycle. Left in constant conditions, the clock will free-run with a period of only approximately 24 hours such that its internal time, or phase, drifts away from that of its environment. Thus, vital to a circadian clock is its ability to entrain to external time through environmental factors (Boulos *et al*., 2002; Dunlap *et al*., 2004; Daan *et al*., 1976).

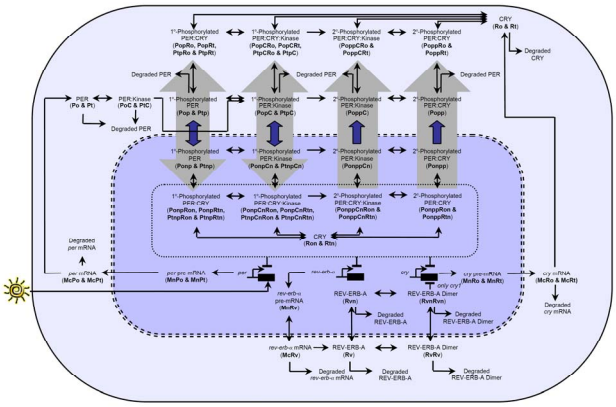

Fig. 1. Gene regulatory network underlying circadian rhythms in neurons in the SCN. The large grey arrows denote activated complexes of proteins that inhibit the transcription of their corresponding genes, thus leading to time-delayed negative feedback and oscillations.

3.2 Apoptosis – Programmed Cell Death

A second example is the apoptosis network in which an extracellular input "controls" the response of the cell as a result of this information processing network. Apoptosis is the "programmed cell death" machinery that is used by nature to strategically kill off un-needed cells, but this mechanism becomes impaired in cancer, leading to unchecked proliferation.

The specific example here is triggered by the ligand Fas. When an activated T-cell contacts a diseased cell, Fas and its natural ligand bind, resulting in the formation of the death inducing signaling complex. This complex then activates two pathways that both lead to the activation of the so-called executioner caspase 3. The topology of the network has been modelled by (Hua *et al*., 2006) and is illustrated in Fig. 2. In the Type I pathway, a feedback involving caspase-6 and caspase-8 regulates the amount of activated caspase-3. In the Type II pathway, Bcl-2 and active caspase-8 both interact with mitochondria, affecting the mitochondrial permeability. Caspase-8 allows mitochondria to become active (permeable), encouraging apoptosome and Smac (second mitochondrial-activator caspase) activation (Stucki *et al*., 2005). The activated apoptosome in turn activates caspase-3, while Smac can remove XIAP (X-linked inhibitor of apoptosis protein), further enhancing caspase-3 activation. FLIP, Bcl-2, and XIAP are antagonists to the apoptotic signal. In Type I activation, significant levels of caspase-8 are required for caspase-3 activation. Yet, in Type II cells, only a small amount of caspase-8 is sufficient to induce apoptosis as the death signal is indirectly amplified by the mitochondrial activity (Bagci *et al*., 2006).

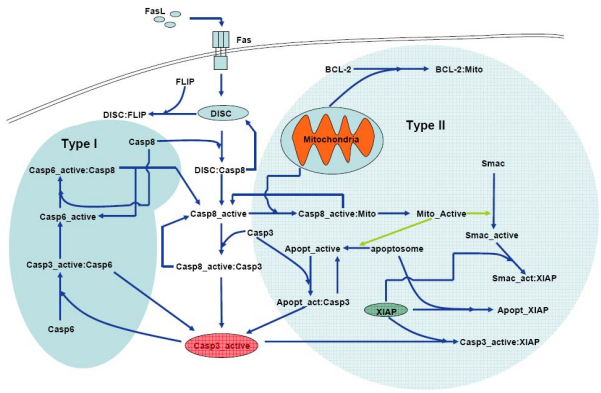

Fig. 2. Network schematic of the Type 1 and Type 2 Fas-induced apoptosis network (adapted from (Hua *et al*., 2006)).

Understanding apoptosis in a broader sense will lead to a better knowledge of the common platform for emergence of cancer cells, and perhaps point to possible cures for certain types of cancer. The complexity of apoptosis, however, makes the understanding very difficult without a systems level approach using a mathematical representation of the pathway. Further, analysis of an apoptosis model can reveal the fragility points in the mechanism of programmed cell death that can have physiological implications not only for explaining the emergence of cancer cells but also for designing drugs or treatment for reinstating apoptosis in these cells.

3.3 Insulin Signaling Pathway

In healthy cells, the uptake of glucose is regulated by insulin, which is secreted by the pancreas. Simply stated, in patients with type 1 diabetes, the pancreas does not produce insulin, whereas in type 2 diabetes, among other consequences, the cells are unable to utilize the insulin produced by the pancreas. The latter phenomenon is best understood from detailed consideration of the insulin signalling pathway (illustrated in Figure 3). The sequence of actions occurs as follows: (i) insulin binds to a receptor on the cell surface, which causes receptor autophosphorylation and activation; (ii) the activated insulin receptor then phosphorylates insulin receptor substrate-1 (IRS1), which subsequently forms a complex with phosphatidylinositol-3-kinase (PI3K); (iii) the IRS1-PI3K complex catalyzes the production of phosphatidylinositol triphosphate (PIP_3), which then interacts allosterically with phosphosinositide-dependent kinase 1 (PDK_1); (iv) the PIP_3-PDK1 complex phosphorylates protein kinases Akt and protein kinase C (PKCζ); (v) activated Akt and PKCζ trigger glucose transporter (GLUT4) translocation from an internal compartment to the cell membrane. In a healthy cell, this cascade ultimately leads to uptake of glucose, and normal "homeostasis". In a cell characterized by type 2 diabetes, the cascade becomes resistant, and the effectiveness of the signal is diminished.

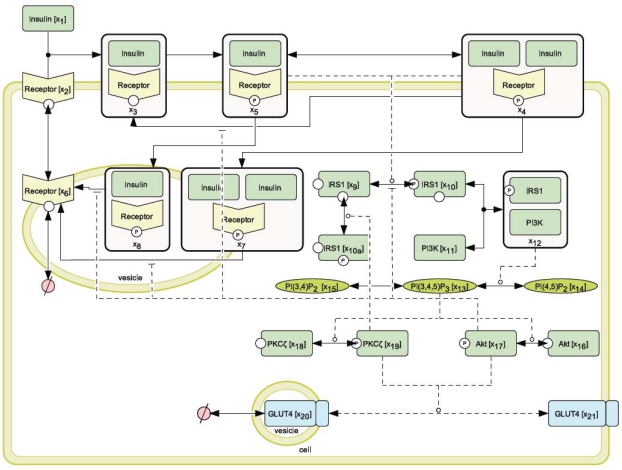

Fig. 3. Insulin signalling pathway model, adapted from (Sedaghat *et al*., 2002).

3.4 Robustness and Bio-inspired Design

Biological networks offer a number of opportunities for inspired design of engineering networks. Aside from the overlapping computational toolkit (e.g., simulation methods for high dimensional, stochastic, stiff, multi-scale systems), there are numerous behaviors in biological networks that offer promise for improved communications and sensors networks. Given the space constraints, we highlight only two of them here, but refer to the reader to the thorough NRC

report on Network Science for additional details (Natl. Acad. Press, Washington DC, 2005).

The coexistence of extreme robustness and fragility constitutes one of the most salient features of highly evolved or designed complexity (Stelling *et al*., 2004b). Optimally robust systems are those that balance their robustness to frequent environmental variations with their coexisting sensitivity to rare events. As a result, robustness and sensitivity analysis are key measures in understanding and controlling system performance. Robust performance reflects a relative insensitivity to perturbations; it is the persistence of a system's characteristic behavior under perturbations or conditions of uncertainty. Measuring the robustness of a system determines the behavior (the output or performance) as a function of the input (the disturbance). Formal sensitivity analysis allows the investigation of robustness and fragility properties of mathematical models, yielding local properties with respect to a particular choice of parameter values.

3.5 Synchronization

Synchronization is manifested in natural circuits in, among other ways, the coherent response of an ensemble of otherwise noisy components. There are a number of examples of this behavior in neuronal systems, most notably the coherence in circadian timekeeping achieved by cells in the suprachiasmatic nucleus (Herzog et al., 2004). The remarkable feature of these cells is that their individual characteristics are remarkably diverse, with individual periods ranging from 21 to 26 hours, and moreover, a given cell may exhibit fluctuations in the period of its firing rate from 2-10 hours between consecutive cycles. Despite this huge range of "component tolerances", the intact signaling network is able to reduce this dispersion by over an order of magnitude in timekeeping precision. By contrast, engineering networks are built from high precision components, yet often struggle with network performance characteristics such as time synchrony. Detailed mathematical models have begun to shed light on the molecular originals of this robustness (To *et al*., 2006).

4. ROBUSTNESS IN THE CIRCADIAN CLOCK

4.1 Sensitivity Analysis for Robustness Analysis

In our work over the last 10 years, we have demonstrated that sensitivity analysis can provide unique insights into the functioning of complex biophysical networks. Of particular interest is the behaviour of biological circuits that exhibit oscillations (e.g., circadian rhythm, cell cycle, neuron firing, cardiac cycles, etc.). Space precludes a detailed mathematical review of the methods employed, instead we will highlight some conclusions of those analyses and refer the interested reader to the original sources for detailed results (Gunawan and Doyle III, 2006; Doyle III *et* al., 2006; Doyle III and Stelling, 2006; Gunawan and Doyle, 2007).

In (Bagheri *et* al., 2007a), we introduced a novel set of sensitivity metrics for performance that were based on a number of different phase-measures: period, phase, corrected phase and relative phase. Our motivation was that phase appears to be the biological imperative, rather than period, for optimal regulation. Both state- and phase-based tools were applied to free-running (absence of light—dark cycles) *Drosophila melanogaster* and *Mus musculus* circadian models. Each metric produced unique sensitivity values used to rank parameters from least to most sensitive. Similarities among the resulting rank distributions strongly suggested a conservation of sensitivity with respect to parameter function and type. A consistent result, for instance, is that model performance of biological oscillators is more sensitive to global parameters than local (i.e. circadian specific) parameters. Differences across the metrics revealed that the conclusions about robustness were dependent on the metric employed for performance.

In (Taylor *et* al., 2008), we derived a novel sensitivity measure, the parametric impulse phase response curve (pIPRC), which both characterizes the phase behavior of an oscillator and provides the means for computing the response to an arbitrary signal (in the form of parametric perturbation). The pIPRC builds on the knowledge that biologists have collected for decades in the form of phase response curves (PRCs), to more general classes of input perturbations. The PRCs and infinitesimal PRCs presented in that study provided quantifiable measures of robustness for oscillators acting as pacemakers. In these systems, robust performance involves proper maintenance of phase behavior. In the case of the circadian clock, this means that the PRC to light must have not only the proper shape, but also the correct magnitude. In previous work (Zeilinger *et al.*, 2006), we were able to invalidate a model of the circadian clock in the plant *Arabidopsis thaliana*, because the pIPRC had neither the proper shape nor the proper magnitude.

In our most recent work (Mirsky *et* al., 2008) we have employed sensitivity methods to predict the likelihood that noise propagates in stochastic models of the circadian network. We found that noise introduced into a sensitive point in the clock propagates very well while noise introduced into an insensitive (or robust) point is undetectable elsewhere. The noise propagates without regard to distance from point of introduction to point of measurement. We conclude that the sensitive global parameters are the sites of effective noise propagation in the clock. We hypothesize that global parameters govern reactions at critical points in our network and therefore may suggest those parts of other systems most worthy of investigation.

The previous studies had focused on single cell models, yet – as the title of this paper suggests – network properties must also be analyzed at higher levels in the system. In (Bagheri *et* al., 2008b), we demonstrated that computational techniques applied to single cell data are fundamental for tuning and predicting the behaviour of oscillatory phenomena at the population level (To *et al*., 2007), since the results of such investigations point to the coupling mechanisms that give rise

to spontaneously synchronized networks of stochastic biophysical nodes. Without such insight, we would not have been able to reproduce the synchrony observed in the SCN. As a result, it is important for experimental biologists to adopt the tools necessary to analyze the structure of both in vitro and in vivo systems.

4.2 Control Insights for Novel Drug Targets

As the previous section reviewed, sensitivity analysis can be used to develop insight on the parts of a network that are most sensitive, consequently, the most susceptible to intervention such as the targeting of a drug. However, a typical application will require a temporal forcing of a node (or nodes) in a network to elicit an optimal response. In our recent work, we have employed model predictive control algorithms to generate the optimal "forcing protocol" that will reset the circadian clock from a condition of phase offset (i.e., jet-lag) (Bagheri *et al*., 2007b; Bagheri *et al*., 2008a). Through parametric state sensitivity analysis, we identified key driving mechanisms for optimal manipulation of the large complex circadian network. We demonstrate, for instance, that the use of non-photic control inputs outperform light-based phase resetting dynamics. Aside from targeting individual parameters as control inputs, our Fisher Information Matrix based parametric sensitivity analyses identified combinations of parameters for control (i.e., vector strategies). The derived MPC algorithm is found to be robust to model mismatch and outperforms the open-loop 24 hour sun cycle based phase recovery strategy by nearly 3-fold.

5. SOME OPEN TECHNICAL CHALLENGES

The field of systems biology is, in many respects, at an early stage – but it is clear that the control and dynamics community have become quickly integrated as an essential component of the interdisciplinary research team that is advancing on multiple frontiers. The open questions are numerous, and we will highlight only a few of them here. A selected list of such challenges was formulated at a recent conference that aimed to bring the control community together with the biologists (Foundations of Systems Biology in Engineering, FOSBE, Santa Barbara, 2005 – www.fosbe.org):

- How can one use systems biology and genomic data to analyze, interpret, and predict the relation between an organism's genotype and its phenotype?
- What are the best methods to combine data from hypothesis-driven research with data from high-throughput studies to create models of cells, communities of cells, and entire organisms? How can one drive iteration and innovation in this model building process?
- What is an appropriate computational infrastructure for maximizing the mining of bioinformatic data? What are the data format and databasing challenges?
- How do the complex network structures constrain intracellular signaling processing?
- What are the promising techniques for estimating model parameters from high throughput data records? How can identification methods be used to drive effective design of experiments?
- How can biological domain knowledge be combined with systems engineering methods to yield model reduction methodologies that capture essential features of biological regulation across multiple scales of time and space?
- How do we educate, train and develop the systems biologists of the future?

We will highlight two of the more critical issues here, that of model inference (model identification), and that of model (in)validation. Clearly, analysis of robustness properties of a biophysical network are predicated upon accurate mathematical descriptions.

5.1 Network Inference and Model Structure

The "inference" problem involves the estimation of the interactions of elements in a biophysical network (e.g., gene-gene, gene-protein, protein-protein, etc.), given time series data of activities of different nodes (e.g., gene interactions from gene expression data). The goals of the inference problem are multiple, and include: (i) hypothesis generation, (ii) design of experiment, (iii) understanding of cellular function, and (iv) unraveling design principles, among others. The sources of information for these inference problems include large scale deletion projects, and vast numbers of microarray experiments. In the early years of bioinformatics studies, the structural localization properties were inferred (e.g., which transcription factors regulate the transcription of which genes), although experimental methods now exist for identifying protein-DNA interactions on a genomic scale, such as ChiP (chromatin immunoprecipitation) assays, that yield structural knowledge.

Given the wide variety of modeling objectives, as well as the heterogeneous sources of data, it is not surprising that many approaches exist for capturing network interactions in the form of mathematical structures, for example:

- Boolean Networks – in which the network is represented as a graph of nodes, with directed edges between nodes and a function for each node (e.g., Ideker *et al*., 2000).
- Petri Nets – another graph theoretic structure in which nodes (or places) are connected by arcs and activities are modeled by transitions (e.g., Nagasaki *et al*., 2004)
- Bayesian Nets – combine directed acyclic graphs with a conditional distribution for each random variable (vertices in graph) (e.g., Pe'er *et al*., 2001)

• Signed Directed Graphs – another graph theoretic structure in which a signed directed edge is used to represent activation versus inhibition (depending on sign) (e.g., Kyoda *et al*., 2004)

• S-systems – notably a dynamic approach in which polynomial nonlinear dynamic nodes are used to capture network behavior (e.g., Kimura *et al*., 2005)

A significant challenge in constructing these network models from data, particularly for gene network models, is the fact that the node dimension (number of genes) can be on the order of 10,000 – leading to a computationally untenable problem for inference (i.e., determination of 108 coefficients of interaction!). In reality the network is tremendously sparse and highly structured, such that there are orders of magnitude fewer "interactions" that must be captured with coefficients. The knowledge that not every gene regulates every other gene, and the fact that not every transcription factor regulates every gene can be exploited to prune significantly the number of coefficients for network identification.

A related concept that can be exploited is the knowledge that the low dimensional connection structures in these networks obey regular hierarchies, which create opportunities for structured model identification. Many biophysical networks 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). 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. 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): (i) feedforward loop, (ii) single input module, and (iii) densely overlapping regulon. Similar studies in a completely different organism, S. cerevisiae, yielded six related or overlapping network motifs (Lee *et al*., 2002): (i) autoregulatory motif, (ii) feedforward loop, (iii) multi-component loop, (iv) regulator chain, (v) single input module, and (vi) multi-input module.

Beyond structural classification, one can analyze these motifs for their functional character, as shown by (Wolf and Arkin, 2003), and again, one finds the recurring dynamic functional motifs in circuits and signal processing: (i) switches, (ii) oscillators, (iii) amplitude filters, (iv) bandpass filters, (v) memory, (vi) noise filters, and (viii) noise amplifiers.

In effect, these studies demonstrate 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 creates opportunity for network inference methods that incorporate such knowledge via constrained search methods, or exploiting prior knowledge in Bayesian frameworks.

In addition to the two classes of models mentioned previously (based on complexity and detail), there is an intermediate class consisting of optimization-based models. In many respects, this class has a hybrid character of empiricism and fundamental details. 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 (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 potential shortcoming is a limited handling of more flexible objective functions that are commonly observed in biological systems. 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 an alternate formulation, dynamic mechanistic details are incorporated as constraints leading to a dynamic FBA extension (Mahadevan, *et al*., 2002).

As we have noted multiple times in this paper – dynamic behavior is an essential property of complex biophysical networks that must be captured in models of those networks. There are some preliminary ideas in capturing network behavior in the form of dynamic models – both discrete time (Hartemink *et al*., 2002) and continuous (Zak *et al*., 2004). There are many challenges in developing dynamic models from the type of data that is typically generated in the corresponding experiments, including: (i) sampling rate is rarely uniform and (ii) data is often the combined with other labs, introducing a number of biases. The previously noted

problems of the curse of dimensionality are more pronounced in the case of dynamic models, if one augments the network interconnection dimensionality with a large number of possible dynamic states (activated, repressed, silenced, etc.), let alone the full continuum of dynamic response.

5.2 Validation, Iteration, Discrimination, and Identifiability

One of the major issues in reverse-engineering of genetic regulatory network is the challenge of uniquely identifying the gene interactions (i.e., model parameters) from experimental data, such as gene expression profiling. This issue, known as identifiability in control theory (Ljung, 1999), deals with the information content of the data; the quantity and quality of the measurements with respect to the model parameters. Recent work in the US and in Europe on the identifiability of gene networks revealed that full knowledge of gene interconnections and perfect measurements still could not guarantee full identifiability of gene interactions (Zak *et al*., 2003), and, furthermore, that improved experimental protocol was far more effective than increased measurements (J. Stelling, unpublished data, 2005). The latter study points to the fact that perturbations should 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 in which a mediating component (e.g., mRNA) is introduced to correct an intermediate element in the direct action cascade described previously.

Coupled to this, noise in the measurements and the inherent stochastic nature of gene expression make practical identification of genetic regulatory networks difficult. In practice, the reverse-engineering of gene network should involve a careful design of the experiments using prior knowledge of the system, to obtain the most informative measurements. Further, this process should be iterative in which the result from each trial is used to better design the next experiment. Here, a measure of the informativeness of data, such as the Fisher Information Matrix (FIM), can lead to a formal procedure for the optimal design of experiment. Aside from the aspect of the quality of data, another practical limitation in most (if not all) of the reverse-engineering of gene network is the limited quantity of data, in terms of sampling frequency and number of independent measurements. For example, although gene expression profiling can provide high throughput data to estimate interactions among thousands of genes, this method still does not depict the protein-mediated regulatory effects. In many cases, parameter estimation from limited measurements suffers from stringent computational requirement and degeneracy, where many parameter combinations give similar agreement to the observed behavior. Here, measurement selection procedures can help identify the combination of measurements that give the best identifiability.

Given the iterative nature of this framework for model development and refinement of experimental protocol, a termination criterion must be established. In the application domain of systems engineering, it is understood that for certain experimental data, it is not possible to confirm whether the model is really valid; however, one can conclude whether the model is not contradicted by the given data (Poolla *et al*., 1994). Such model (in)validation tests can be formulated for the network inference problems described in this chapter, and are usually based on the difference between the simulated and measured output and some statistics about these differences. Typical statistics for the model errors include maximum absolute value, mean value and variance. These methods are slowly migrating from the engineering domain, and are likely to find greater application in systems biology as experimental methods are refined, and closer collaborations are developed between modelers and experimentalists.

8. SUMMARY

One might conclude from all of this that the aims of control researchers in the field of systems biology are to elucidate mechanisms in biology to advance the state of understanding. While this is certainly true, there is much broader impact of this line of investigation. In particular, medical therapeutics and treatments are the focus of many of the research teams working in systems biology.

The medical driving forces include the identification of "targets" in the network for therapeutic intervention. A "systems" analysis reveals that a single point perturbation is often less effective than a vectoral perturbation, and at the same time, a point perturbation will often propagate beyond the intended action, leading to undesirable side effects. Systems methods are also applied to these networks to determine "signatures" of the propagation of a disease state (or markers). For example, the temporal progression of the apoptic response in the network in Figure 2 could be tracked by gene expression profiling, and one can determine the time course of the response, and apply appropriate therapy at the optimal point in the disease progression.

The size and complexity of cellular networks make intuition inadequate for deducing cellular behavior from the underlying gene and protein interactions. Such analysis is critical to guide the development of medical solutions for problems where the network has "failed". In cellular networks, high sensitivities or strong gain directions point to the weakest links in the system. Perturbations on these links can potentially lead to a large disruption in the network behavior, i.e., the network is not robust (fragile) to the uncertainty in these pathways. These "hot spots" (or fragile nodes) have several implications. First, further model refinements on the hot spots in the network may be necessary. Second, when the model is sufficiently accurate, these "targets" offer coherent strategies for intervention in an otherwise complex circuit diagram (such as those depicted in Figures 1, 2, and 3). In the case of circadian timekeeping, the regulatory network insights could shed light on drug targets for jet lag or sleep disorders. Of course, one has to link the subcellular network in Figure 1 with whole body function (Kronauer *et al.*, 2007). Tracking the progression of apoptosis through the network in Figure 2 can lead to both

signatures of the disease for monitoring, as well as vectoral intervention strategies for treatment. For the insulin signalling pathway (Figure 3), opportunities to restore insulin sensitivity could result in the case where the pathway has become resistant (type 2 diabetes).

It is indeed an early stage in the field of systems biology, with such complex biophysical networks waiting to be unravelled, but it is abundantly clear that control and dynamics researchers will play a key role as this field progresses.

ACKNOWLEDGMENTS

The ideas in this paper and the corresponding plenary lecture are the product of many, many fruitful discussions and research projects with my students, post-docs, and collaborators. The research ideas should be credited to them: Neda Bagheri, Erik Herzog, Steve Kay, Eric Kwei, Doug Lauffenburger, Kirsten Meeker, Henry Mirsky, Linda Petzold, Jason Shoemaker, Joerg Stelling, Stephanie Taylor, and Melanie Zeilinger.

This work was supported by the Institute for Collaborative Biotechnologies under Grant DAAD19-03-D-0004 from the U.S. Army Research Office, IGERT National Science Foundation (NSF) under Grant DGE02-21715, NSF/NIGMS Grant GM078993, and Army Research Office under Grant W911NF-07-1-0279.

REFERENCES

Angeli, D., J. E. Ferrell and E. D. Sontag (2004). Detection of multistability, bifurcations, and hysteresis in a large class of biological positive-feed back systems. *Proceedings of the National Academy of Sciences of the United States of America,* **101**, 1822-1827.

Arkin, A., J. Ross and H. H. McAdams (1998). Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected Escherichia coli cells. *Genetics,* **149**, 1633-1648.

Bagci, E. Z., Y. Vodovotz, T.R. Billiar, G.B. Ermentrout, and I. Bahar (2006). Bistability in apoptosis: Roles of bax, bcl-2, and mitochondrial permeability transition pores. *Biophys. J*. **90**, 1546–1559.

Bagheri, N., J. Stelling and F.J. Doyle III (2007a). Quantitative Performance Metrics for Robustness in Circadian Rhythms. *Bioinformatics*, **23**, 358-364.

Bagheri, N., J. Stelling and F.J. Doyle III (2007b). Circadian Phase Entrainment via Nonlinear Model Predictive Control. *Intl. J. Robust and Nonlinear Control*, **17**, 1555-1571.

Bagheri, N., J. Stelling, and F.J. Doyle III (2008a). Circadian Phase Resetting & Multiple Control Targets. *PLoS Comp. Biol*., accepted.

Bagheri, N., S.R. Taylor, K. Meeker, L.R. Petzold, and F.J. Doyle III (2008b). Synchrony and Entrainment Properties of Robust Circadian Oscillators. *Roy. Soc. Interface*, accepted.

Barabasi, A.L. (2004). Network Biology: Understanding the Cell's Functional Organization. *Nature Rev. Genetics*, **5**, 101-113.

Barkai, N. and S. Leibler (1997). Robustness in simple biochemical networks. *Nature,* **387**, 913-917.

Barkai, N. and S. Leibler, (2000). Circadian Clocks Limited by Noise. *Nature,* **403**, 267–8.

Boulos, Z., Macchi, M.M., Sturchler, M.P., Stewart, K.T., Brainard, G.C., Suhner, A., Wallace, G., and R. Steffen (2002). Light visor treatment for jet lag after westward travel across six time zones. *Aviat. Space Environ. Med*., **73**, 953-963.

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.

Daan, S., and C.S. Pittendrigh (1976). A functional analysis of circadian pacemakers in nocturnal rodents. II. The variability of phase response curves. *J. Comp. Physiol*., **106**, 253-266.

Doyle, F.J. III, R. Gunawan, N. Bagheri, H. Mirsky, and T.L. To (2006). Circadin Rhythm: A Natural Robust, Multi-Scale Control System. *Comp. Chem. Eng*., **30**, 1700-1711.

Doyle, F.J. III and J. Stelling (2006). Systems Interface Biology. *J. Royal Soc. In*terface, **3**, 603-616.

Dunlap, J.C., Loros, J.J., and P.J. DeCoursey (Eds.) (2004). *Chronobiology: Biological Timekeeping*. Sinauer Associates, Inc., Sunderland, MA.

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,

Goldbeter, A. (1996). *Biochemical oscillations and cellular rhythms: the molecular bases of periodic and chaotic behavior*. Cambridge, Cambridge University Press.

Gunawan, R. and F.J. Doyle III (2006). Isochron-based Phase Response Analysis of Circadian Rhythms. *Biophys. J*., **91**, 2131-2141.

Gunawan R. and F.J. Doyle III (2007). Phase Sensitivity Analysis of Circadian Rhythm Entrainment. *J. Biol. Rhythms*, **22**, 180-194.

Hartemink, A.J., D.K. Gifford, and T.S. Jaakola (2002). Combining Location and Expression Data for Principled

Discovery of Genetic Regulatory Network Models. *Proc. Pac Symp Biocomput*, **7**, 437–449.

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

Herzog, E.D., Aton, S.J., Numano, R., Sakaki, Y., and H. Tei (2004). Temporal precision in the mammalian circadian system: a reliable clock from less reliable neurons. *J. Biol. Rhythm*, **19**, 35-46.

Hua, F., S. Hautaniemi, R. Yokoo, and D.A. Lauffenburger (2006). Integrated mechanistic and data-driven modelling for multivariate analysis of signalling pathways. *J. R. Soc. Interface,* **9**, 515–26.

Ideker, T., T. Galitski and L. Hood (2001). A new approach to decoding life: Systems biology. *Annual Review of Genomics and Human Genetics,* **2**, 343-372.

Ideker, T., V. Thorsson, and R.M. Karp (2000). Discovery of Regulatory Interactions Through Perturbations: Inference and Experimental Design. *Pac. Symp. BioComputing*,

Kimura, S., K. Ide, A. Kashihara, M. Kano, M. Hatakeyama, R. Masui, N. Nakagawa, S. Yokoyama, S. Kuramitsu, and A. Konagaya (2005). Inference of S-system Models of Genetic Networks using a Cooperative Coevolutionary Algorithm. *Bioinformatics*, **21**, 1154-1163.

Kitano, H. (2001). *Foundations of Systems Biology*, MIT Press.

Kitano, H. (2002). Systems biology: a brief overview. *Science*, **295**, 1662-1664.

Klipp, E., R. Herwig, A. Kowald, C. Wierling and H. Lehrach (2005). *Systems Biology in practice: concepts, implementation and application*. Weinheim, Germany, Wiley.

Kronauer, R., G., Gunzelmann, H. Van Dongen, F.J. Doyle III, and E. Klerman (2007). Modeling Human Neurobehavioral Performance I: Uncovering Physiologic Mechanisms*. J. Biol. Rhythms*, **23**, 233-245.

Kyoda, K., K. Baba, S. Onami, and H, Kitano (2004). DBRF-MEGN Method: An Algorithm for Deducing Minimum Equivalent Gene Networks from Large-Scale Gene Expression Profiles of Gene Deletion Mutants. *Bioinformatics*, **20**, 2662-2675.

Lee, T.I., N.J. Rinaldi, F. Robert, D.T. Odom, Z. bar-Jospeh, 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.

Liu, A.C., Welsh, D.K., Ko, C.H., Tran, H.G., Zhang, E.E., Priest, A.A., Buhr, E.D., Singer, O., Meeker, K., Verma, I.M., Doyle III, F.J., Takahashi, J.S., and S.A. Kay (2007). Intercellular coupling confers robustness against mutations in the SCN circadian clock network. *Cell*, **129**, 605-616.

Ljung, L. (1999). *System Identification: Theory for the User*. Prentice Hall. Englewood Cliffs, NJ.

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

Malcolm, A.M., and L.J. Heyer (2003). In: *Discovering Genomics, Proteomics, and Bioinformatics*. (Benjamin Cummings, San Francisco.

Mirsky, H., R. Gunawan, S. Taylor, J. Stelling, and F.J. Doyle III (2008). Noise Propagation and Sensitivity in Mammalian Circadian Clocks. *PLoS Comp. Biol*., submitted.

Muller, T. G., D. Faller, J. Timmer, I. Swameye, O. Sandra and U. Klingmuller (2004). Tests for cycling in a signalling pathway. *Journal of the Royal Statistical Society Series C-Applied Statistics,* **53**, 557-568.

Nagasaki, M., A. Doi, H. Matsuno, and S. Miyano (2004). A Versatile Petri Net Based Architecture for Modeling and Simulation of Complex Biological Processes. *Genome Informatics*, **15**, 180-197.

National Research Council: Network Science. (National Academies Press, Washington DC, 2005)

Palsson, B. (2006). *Systems biology: properties of reconstructed networks*. Cambridge, UK, Cambridge University Press.

Pe'er, D. A. Regev, G. Elidan, and N. Friedman (2001). Inferring Subnetworks from Perturbed Expression Profiles. *Bioinformatics*, **17**, S215-S224.

Poolla, K., P. Khargonekar, A. Tikku, J. Krause, and K. Nagpal (1994). A Time-Domain Approach to Model Validation. *IEEE Trans Automat Contr*, **39**, 951–959.

Rao, C. V., M. Frenklach and A. P. Arkin (2004). An allosteric model for transmembrane signaling in bacterial chemotaxis. *J Mol Biol*, **343**, 291-303.

Reppert, S.M. and D.R. Weaver (2002). Coordination of circadian timing in mammals. *Nature,* **418**, 935-941

Sedaghat, A.R., A. Sherman, and M. J. Quon (2002). A mathematical model of metabolic insulin signaling pathways. *American Journal of Physiology-Endocrinology and Metabolism*, **283**, E1084-E1101.

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.

Sontag, E. D. (2002). Asymptotic amplitudes and Cauchy gains: a small-gain principle and an application to inhibitory biological feedback. *Systems & Control Letters*, **47**, 167-179.

Stelling, J. (2004). Mathematical Issues in Systems Biology. *Current Opinion in Microbiology*, **7**, 513-518.

Stelling, J., E. D. Gilles and F. J. Doyle (2004a). Robustness properties of circadian clock architectures. *Proceedings of the National Academy of Sciences of the United States of America,* **101**(36), 13210-13215.

Stelling, J., U. Sauer, Z. Szallasi, F. J. Doyle and J. Doyle (2004b). Robustness of cellular functions. *Cell,* **118**, 675-685.

Stucki, J. W. and H.-U. Simon (2005). Mathematical modeling of the regulation of caspase-3 activation and degradation. *J. theor. Biol*. **234**, 123–131.

Szallasi, Z., J. Stelling and V. Periwal, Eds. (2006). *System modeling in cellular biology: from concepts to nuts and bolts*. Cambridge, MA, MIT Press.

Taylor, S., L. Petzold, and F.J. Doyle III (2008). Sensitivity Measures for Oscillating Systems: Application to Mammalian Circadian Gene Network. *IEEE Trans. Aut. Control*, **53**, 177-1888.

To, T.L., M.A. Henson, E.D. Herzog, and F.J. Doyle III (2007). A Molecular Model for Intercellular Synchronization in the Mammalian Circadian Clock. *Biophys. J*., **92**, 3792-3803.

Varma, A., M. Morbidelli, and H. Wu (1999). *Parametric Sensitivity in Chemical Systems*. Oxford University Press, New York, NY.

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

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

Wolf D.M. and A.P. Arkin (2003). Motifs, Modules, and Games in Bacteria. *Current Opinion in Microbiology*, **6**, 125-134.

Yi, T. M., Y. Huang, M. I. Simon and J. Doyle (2000). Robust perfect adaptation in bacterial chemotaxis through integral feedback control. *Proceedings of the National Academy of Sciences of the United States of America*, **97**, 4649-4653.

Zak, D., G. Gonye, J.S. Schwaber and F.J. Doyle III (2003). 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*, **13**, 2396–2405.

Zak, D., R.K. Pearson, R. Vadigepalli, G. Gonye, J.S. Schwaber and F.J. Doyle III (2004). Continuous-time identification of gene expression models. *Omics*, **7**, 373-386.

Zeilinger, M.N., E.M. Farre, S.R. Taylor, S.A. Kay, and F.J. Doyle III (2006). A novel computational model of the circadian clock in Arabidopsis that incorporates PRR7 and PRR9. *Nature Molecular Systems Biology*, **2**, 1-13.