Logic-Based Models for the Analysis of Cell Signaling Networks[†]

Melody K. Morris,[‡] Julio Saez-Rodriguez,[‡,§] Peter K. Sorger,[‡,§] and Douglas A. Lauffenburger*,[‡]

[‡]*Center for Cell Decision Process and Department of Biological Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and* [§]*Department of Systems Biology, Harvard Medical School, Boston, Massachusetts 02115*

ABSTRACT: Computational models are increasingly used to analyze the operation of complex biochemical networks, including those involved in cell signaling networks. Here we review recent advances in applying logic-based modeling to mammalian cell biology. Logic-based models represent biomolecular networks in a simple and intuitive manner without describing the detailed biochemistry of each interaction. A brief description of several logic-based modeling methods is followed by six case studies that demonstrate biological questions recently addressed using logic-based models and point to potential advances in model formalisms and training procedures that promise to enhance the utility of logic-based methods for studying the relationship between environmental inputs and phenotypic or signaling state outputs of complex signaling networks.

With accelerating pace, molecular biology and biochemistry are identifying complex patterns of interactions among intracellular and extracellular biomolecules. With respect to cell signaling in eukaryotes, the focus of this review, complex multicomponent networks involving many shared components govern how a cell will respond to diverse environmental cues. Powerful experimental approaches now exist for identifying components of these networks and for determining their biochemical activities, but understanding the networks as an integrated whole is difficult using intuition alone. Thus, mathematical and computational modeling is increasingly playing a role in data interpretation and attempts to extract general biological understanding (*1*, *2*). Depending on the network studied, the data available, and the questions posed, a diverse spectrum of modeling approaches exists, ranging from the highly abstract to the highly specified (*3*, *4*). The goal of this review is to discuss logic-based modeling, an approach lying midway between the complexity and precision of differential equations on one hand and data-driven regression approaches on the other.

Within the spectrum of modeling methods currently being applied to cellular biochemistry, models involving differential equations bear the closest relationship to underlying biochemical rate laws. Sets of coupled ordinary differential equations (ODEs) can effectively represent chemical reactions when the number of molecules is large and mass action approximations are appropriate. Partial differential equations (PDEs) add the ability to represent spatial gradients (*5*), and stochastic methods make it possible to analyze systems in which the number of molecules is small (*6*). Networks of differential equations can model the temporal and spatial dynamics of biochemical processes in considerable detail, making it possible to study chemical mechanism and predict network dynamics under various conditions. However, the topology of ODE- and PDE-based models (that is,

patterns of interaction among the species) must be specified in advance, and model output is strongly dependent on the values of free parameters (typically initial protein concentrations and rate constants). Estimating these parameters is a computationally intensive task requiring substantial data. As networks get larger, ODE modeling becomes more and more challenging, and models that attempt to capture real biological data are currently limited to a few dozen components.

At the other extreme, a very active field for computing graphical representations of biological networks through literature analysis or identification of correlations in high-throughput data has emerged. In these graphs, termed protein interaction networks (PINs or interactomes) or protein signaling networks (PSNs), genes and proteins are represented by nodes and potential interactions by edges (links). The edges can be directional or not and signed (inhibitory/activating) or not and typically represent a wide range of interaction modes from direct physical binding to correlated gene expression (*7*) or integrated database entries (*8*). Graphs are an attractive way to summarize diverse relationships among large numbers of biomolecules across multiple organisms, but they are not executable per se and cannot be used to compute input−output relationships. Moreover, network graphs rarely take into account dynamic changes in signaling activities, cell type-specific biochemistry, or context-dependent variations (*9*).

Here, we review logic-based models, which represent a compromise between highly specified differential equation models and protein interaction graphs. Using logic-based methods, it is possible to model interactions among large numbers of protein species and perform model training, model validation, and model-based prediction. The first application of logic-based modeling to biological pathways is credited to Kauffman, who used discrete logic to model the biological process of gene regulation (*10*). Subsequent work focused on delineating theoretical properties of logic-based models of gene regulation (*11*, *12*). Huang and Ingber were among the first to apply logic-based modeling to cell signaling networks, demonstrating that specific cell phenotypes might correspond to dynamic steady states of a

logic-based model of intracellular signaling species (*13*). This example of linking environmental inputs to phenotypic outputs via a logic-based model of a biochemical signaling network has sparked considerable interest in the possibility of harnessing logic-based models to understand the relationship between biochemical signaling network and cell state, reflected in a large number of studies over the past few years (*13*−*33*).

This review is divided into two sections. In the first, we describe the fundamentals of logic-based modeling; in the second, we discuss six applications of logic-based modeling to eukaryotic biology. We focus on logic-based models of biochemical signaling networks and refer the reader to the literature for a more in-depth explanation of theoretical considerations (*34*), applications of logic-based models to gene regulatory networks (*11*), and models of intercellular communication (*35*, *36*).

REPRESENTING BIOCHEMICAL NETWORKS WITH LOGIC-BASED MODELS

What Is a Logic-Based Model? Consider the graphical representation of a signaling network common to protein inter-action networks (Figure 1a): the nodes in the graph represent proteins, and the edges represent interactions. Such a graph depicts nodes that interact physically or have correlated expression or genetic profiles (depending on the underlying data source) but do not allow us to explicitly compute the state of activity of individual nodes given different inputs or initial network states. Performing such a calculation requires information about how each node reacts to the activities of its input nodes. In logic-based models, these dependencies are specified by "gates" (Figure 1b) which, in Boolean logic, are specified by "truth tables" that list output states for all possible combinations of input states (Figure 1c). Figure 1d shows the truth tables of the OR, AND, and NOT Boolean logic gates as well as a small network in which gates are assembled to create the AND-NOT logic gate.

To illustrate how logic-based modeling can be applied to a biological network, consider a hypothetical representation of epidermal growth factor receptor (EGFR)[1] and several downstream proteins (Figure 1e). This toy network is too simple to be realistic but demonstrates several issues of importance when building a logic-based model. Either epidermal growth factor (EGF) or heregulin (HRG) can bind to and activate EGFR (Figure 1d,e). EGFR then stimulates the Raf/ERK and PI3K/AKT pathways (the multitude of known biochemical interactions in this case are modeled as a single "activating" edge). ERK activity inhibits EGFR-dependent PI3K activation, whereas AKT positively regulates the Raf/ERK pathway (Figure 1d,e). With this information, it is possible to compute the response of the unperturbed network to a given input as well as responses resulting from inhibition of a node (by a drug for example). However, under all simulated conditions [EGF or HRG alone or in combination (Figure 1f)], the network response is the same. This is to be expected because binary logic cannot encode the

differential sensitivities of EGFR to EGF and HRG, a point to which we return below.

Modeling Nondiscrete Processes Using Logic-Based Approaches. The assumption in Boolean logic that all species are either on or off (state 1 or 0, respectively) is clearly an unrealistic way to represent binding curves or catalytic reactions. Fortunately, logic-based modeling provides several approaches for modeling intermediate states of activity (Figure 2a). Multistate discrete models specify additional levels between 0 and 1, whereas fuzzy logic allows for continuous node states. In fuzzy logic, which has found wide utility in industrial control systems, a set of user-defined functions transforms discrete logic statements into relationships between continuous inputs and output levels. Other methods of describing discrete or Boolean logic models as continuous or mixed discrete continuous have also been implemented successfully [Figure 2a (dashed lines)] (*28*, *37*, *38*).

How is a prototypical biological interaction approximated using discrete and nondiscrete logic formalisms? In Figure 2b, a sigmoidal relationship between input and output level [e.g., a protein kinase acting on a substrate (black solid line)] is approximated by binary (red solid line), ternary (green dashed line), and quaternary (blue dashed−dotted line) discrete logic functions. Fuzzy logic and mixed discrete continuous logic can closely approximate the real response (orange dashed line). It is important to note, however, that the increased degree of realism of multistate or fuzzy logic modeling comes at the cost of increased complexity, typically in the form of a threshold or transfer function having free parameters that must be estimated.

Figure 1g provides an example of how multistate discrete logic can be used to represent the differing states of activation of EGFR when exposed to EGF and HRG stimulation, where an additional activation level of "two" indicates that EGFR is more sensitive to EGF than HRG. In the model, addition of HRG alone causes AKT and ERK activity levels to oscillate (Figure 1h, right panel). These oscillations are caused by the negative feedback between ERK and PI3K. However, when either EGF alone or both EGF and HRG are present (Figure 1h, left panel), EGFR is in activation state two and the negative feedback inhibiting PI3K is absent. Thus, oscillations are not observed.

Treatment of Time in a Logic-Based Model. The presence of oscillations in this and other logic-based networks complicates analysis, and the actual form that the oscillations take depends on the treatment of time during the simulation. Logic-based models represent time with varying degrees of detail. We present this concept graphically in Figure 2c, where each modeling formalism is classified according to the detail in its representation in species' state and time. Table 1 presents a comparison of the approaches in tabular form. The activity of each species in discrete logic-based network simulations is determined by its input node states at some previous time step. The order in which node states are updated results in an implicit treatment of time scales.

Two primary node-updating schemes exist: synchronous and asynchronous (*12*, *39*, *40*). Synchronous updating updates every node at each time step according to the states of its input nodes at the previous time step, whereas asynchronous updating updates node states in random order. In practical terms, asynchronous updating involves updating an output node on the basis of some of its input nodes at the current and others at a previous time step. Variants of both synchronous and asynchronous updating exist. Time delays can be specified with synchronous updating, allowing for a more refined description of dynamics. A variant of

---

[1]Abbreviations: EGF, epidermal growth factor; EGFR, epidermal growth factor receptor; ErbB, epidermal growth factor receptor; ERK, extracellular regulated kinase; GAP, GTPase activating protein; HRG, heregulin; IGF1, insulin-like growth factor 1; IKK, IκB kinase; IL1α, interleukin 1α; IL6, interleukin 6; LPS, lipopolysaccharide; MEK, mitogen-activated protein kinase kinase; MK2, mitogen-activated protein kinase-activated protein kinase 2; PBMC, peripheral blood mononuclear cells; PI3K, phosphoinositide 3-kinase; T-LGL, T cell large granular lymphocyte(s); TGFα, transforming growth factor α; Th Cell, T helper cell; TNFα, tumor necrosis factor α.
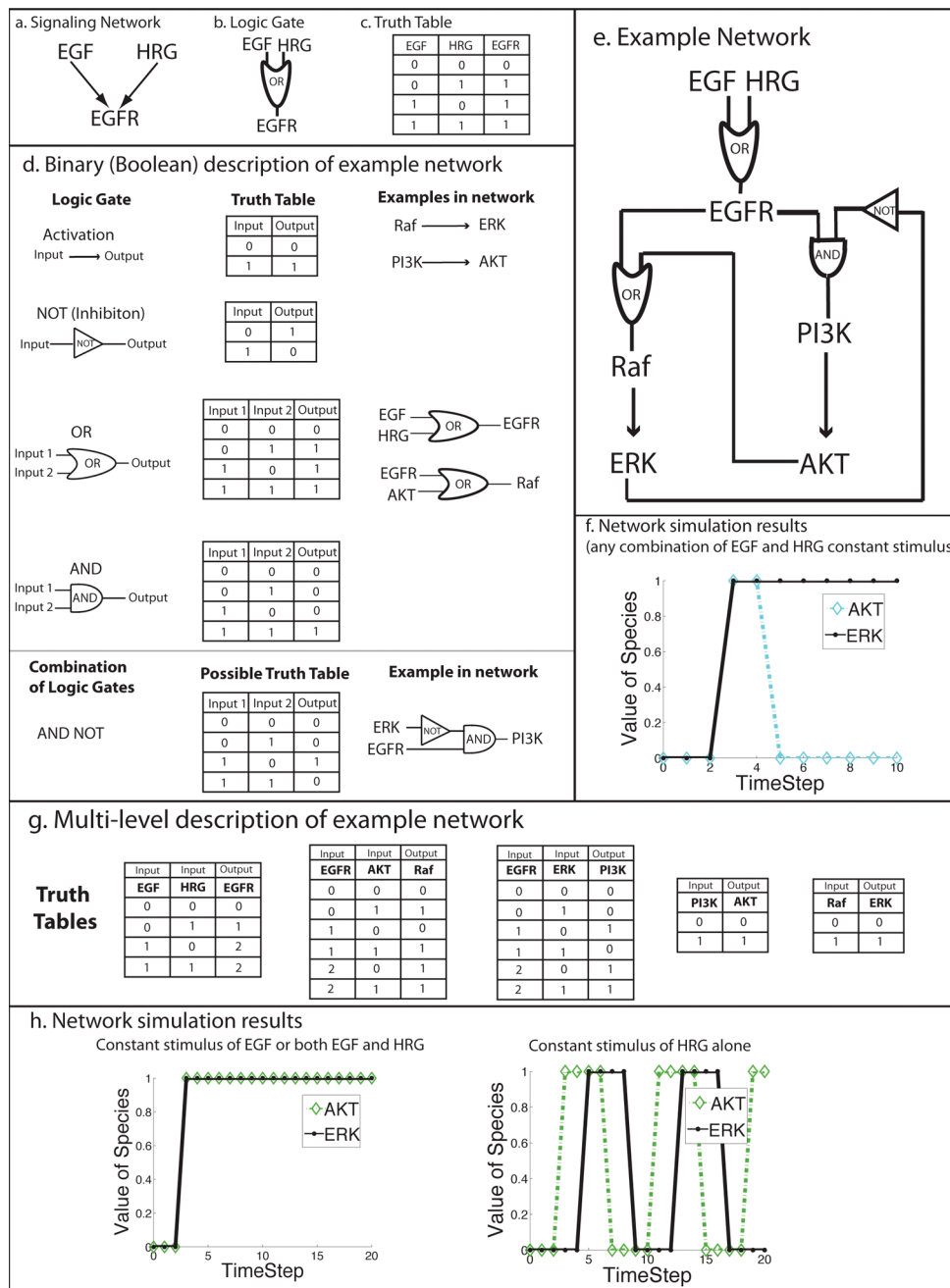
FIGURE 1: Examples of a logic-based network. (a) Protein signaling network. Biochemical species are represented as nodes. The interactions between these nodes are indicated with arrows. (b) Logic gate. Precisely how the nodes interact is specified with a simple Boolean logic gate. (c) Truth table specifying the output node given possible combinations of its inputs nodes' values. (d) Boolean logic gates and their truth tables. If the gates are used in the example network, the interaction is shown on the right. We also describe the AND-NOT gate, which is used in the example network. We note that, in many applications of logic-based modeling, OR and AND gates are not explicitly indicated with their gate symbols. (e) Example of a logic-based network structure. The model was simulated with synchronous updating using custom MatLab (Mathworks, Inc.) code (available as Supporting Information). (f) Network behavior with binary rules. Under initial conditions with different ligand stimulations, the network response was identical because the logic rules did not distinguish between EGF and HRG stimulation. (g) Multistate rule specification. The truth tables are given for each modeled species. These rules specify multiple states. The greater sensitivity of EGFR for EGF than HRG is encoded in the higher level it reaches upon stimulation by EGF. Rules that are different from the binary rules are highlighted. (h) Network behavior with multistate rules given in panel d. The rules specified that EGFR is more sensitive to EGF than HRG. Thus, the behavior differed depending on the stimulation condition. Under EGF or EGF and HRG stimulation, the states of ERK and AKT were stabilized whereas they oscillated under HRG stimulation alone. This is because the rules specified that, with the highest level of activation of EGFR (activation state two), the negative feedback by ERK did not effectively inhibit PI3K, whereas with medium-level activation of EGFR (activation state one accessed with only HRG was present), the negative feedback was effective.

asynchronous updating, mixed asynchronous updating, allows some nodes be updated before others, making it possible to separate time scales of fast (e.g., binding and phosphorylation) and slow (e.g., degradation and transcription) reactions in a manner similar to that of time delays (*41*). Regardless of the

updating scheme, it is frequently observed that logic-based models will settle into an "attractor state" in which states no longer change (logic steady state) or states cycle in a pattern of activity [the oscillations in the example network are an example of a cyclic attractor state (Figure 1h)]. The continuous or mixed
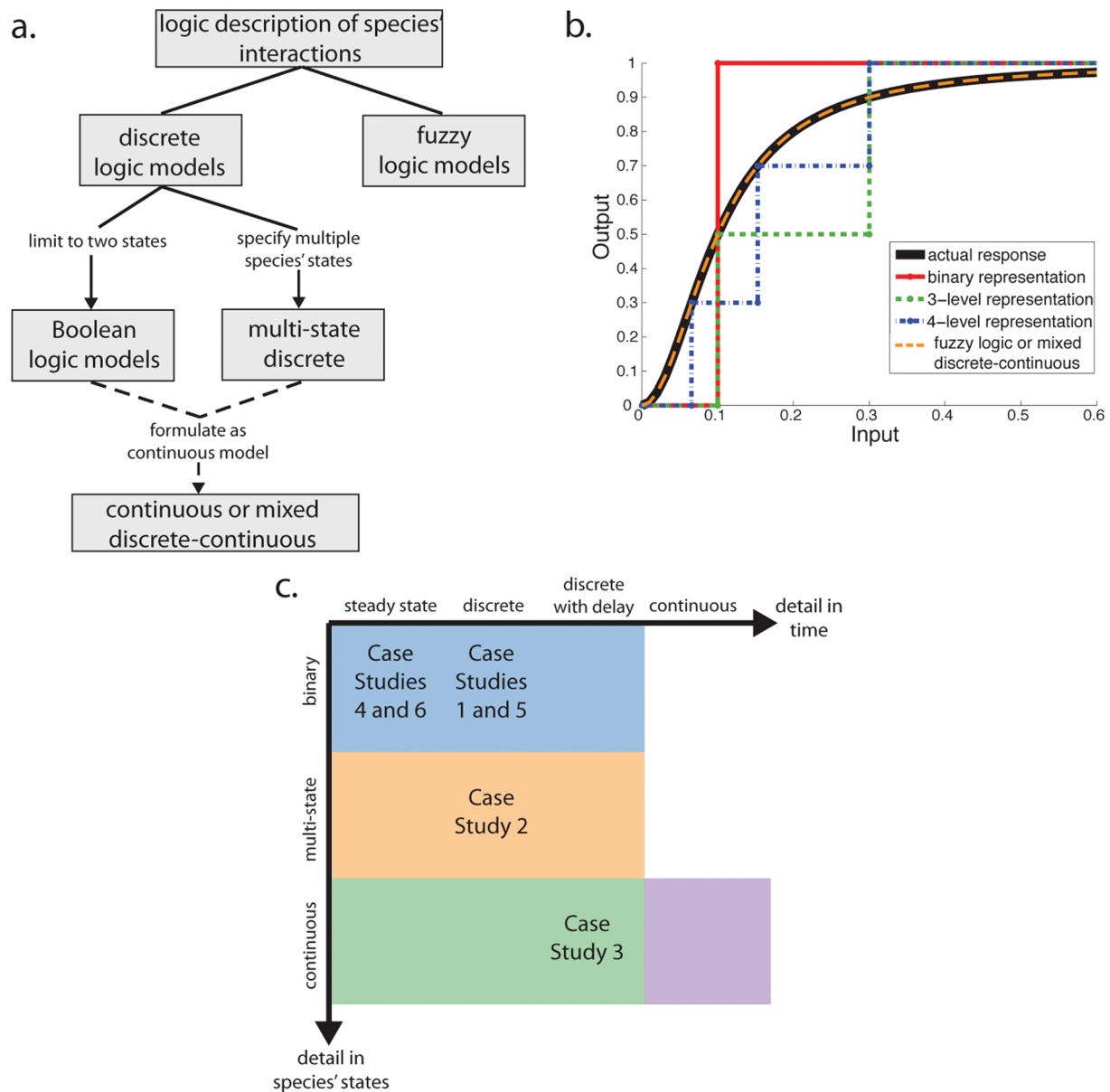
FIGURE 2: Description of logic-based formalisms. (a) Description of various forms of logic-based models. All logic-based models describe species' interactions in terms of logical statements (or rules). Discrete logic can specify two or more levels for each modeled species, whereas Boolean logic specifies only two levels of each species. In addition to these logic-based formalisms, various methods of describing discrete or Boolean logic models with piece-wise continuous equations (*37*) or logic-based ODEs (*28*) have been successfully implemented to represent biochemical signaling networks. (b) Approximation of the input−output relationship between hypothetical biological species (black solid line) with binary (red solid line), ternary (green dashed line), and quaternary (blue dashed−dotted line) discrete logic gates as well as fuzzy logic or mixed discrete-continuous formalisms (orange dashed line). Various thresholds could be chosen for each discrete gate; chosen thresholds are purely hypothetical. (c) Plane of granularity in species' states and treatment of time. Regions containing various logic-based modeling variants are denoted by shaded boxes. Boolean networks (blue region) are binary, but their treatment of time ranges from logic steady state to discrete with delays. Discrete models with multiple species states (orange region) cover a similar range of possible treatments of time. Fuzzy logic models (green region) describe a continuous range of species' states with the same range of time granularity. Conversion of Boolean or discrete models into logic-based ODEs, piecewise linear, and standardized qualitative dynamical system (purple region) results in models that are continuous in both species' states and time. Each case study is placed on the landscape according to how it represents the biological system of interest with a logic-based network.

discrete continuous methods mentioned previously formulate discrete logic as ordinary differential equations or piecewise linear equations, respectively. This treatment allows one to model both species' state and time as continuous (Figure 2c) but at the cost of increased model complexity. Research into the influence of updating scheme on the segment polarity network of *Drosophila melanogaster* (*42*) and the mammalian cell cycle (*43*) network has demonstrated that the different treatments of time can lead to unique biological interpretations. Generally, the most appropriate updating scheme is dependent on the type of model built as well as the questions that the model is meant to address.

Another extension of logic-based modeling aims to incorporate probabilistic interactions (*44*, *45*). This method allows one to account for uncertainty in the knowledge of signaling networks as well as stochasticity in biological systems. Also noteworthy are a number of efforts to apply related formalisms such as Petri nets, cellular automata, etc., to biological networks (*46*). In some cases, these formalisms can be reduced to logic-based formalisms, and they provide an additional level of abstraction that makes it possible to perform formal network analysis (*47*). Because these probabilistic and computational techniques involve slightly different considerations compared to what was previously

Table 1: Description of Logic Modeling Variants

logic modeling variant	time treatment[a]	detail of species' states	use in biological modeling
Boolean	discrete time steps	binary	(13, 15−17, 19, 24, 25, 33);[b] (20, 42, 43, 51, 52)[c]
discrete logic	discrete time steps	multistate; user-defined	(21, 31, 32, 38);[b] (31, 53, 54)[c]
fuzzy logic	discrete time steps or time can be treated as a variable	multistate; user-defined and implicit in calculation of output state	(14, 18, 30)[b]
piecewise linear	continuous	multistate; user-defined and implicit in equations	(42, 55)[c]
logic-based ODEs	continuous	multistate; implicit in ODE equations	28[b]
standardized qualitative dynamical systems	continuous	multistate; implicit in formalism	23[b]

[a]Discrete time steps could use synchronous or asynchronous updating with or without delay or be examined at steady state. [b]Biochemical signaling network. [c]Genetic network.

Table 2: Tools Available for the Logic Modeling of Biochemical Signaling Networks

tool	type of logic	functionality	treatment of time	refs
BooleanNet	Boolean, piecewise linear	simulation and visualization	synchronous, asynchronous, or continuous	56
GinSim	discrete (multistate)	model building, simulation, and analysis	synchronous, asynchronous, or mixed asynchronous	(57, 58)
CellNetAnalyzer	Boolean (multistate)	model simulation, visualization, and network properties analysis	logic steady state	(41, 50)
CellNetOptimizer	Boolean	model training and simulation	logic steady state	25
Odefy	Boolean, logic-based ODEs	model simulation and visualization	synchronous, asynchronous, or continuous	28
Genetic Network Analyzer	piecewise linear	model building and simulation	continuous	59
ChemChains	Boolean	model simulation, visualization, and analysis	synchronous or asynchronous	60
MetaReg	discrete (multistate)	model simulation, visualization, and refinement	logic steady state	61
SQUAD	standardized qualitative dynamical systems	model simulation and analysis	continuous	(38, 62)

discussed, we do not describe them further and instead point the interested reader to the references listed above.

This review focuses on a qualitative description of various logic-based formalisms. For readers interested in the actual computational procedures involved in using these methods, an outline is provided as Supplemental Figure 1 (Supporting Information). Additionally, several dedicated software packages have been developed for logic-based modeling of biochemical signaling networks with varying degrees of detail and differing updating schemes; some of these are listed in Table 2. We refer the interested reader to the references in this table for descriptions of each simulation procedure, in particular the quantitative approaches not described here.

CASE STUDIES OF APPLICATIONS OF LOGIC-BASED MODELS TO BIOCHEMICAL NETWORKS

Below we discuss six logic-based models of signal transduction as a means of highlighting different methods, biological questions, and opportunities for future development; we necessarily omit many details. Figure 3a shows a general workflow for applying logic-based modeling to signaling networks and serves as a means of summarizing the key features of each case study. (i) Case studies 1 and 2 involve models built solely from literature-based prior knowledge (Figure 3b). (ii) Case study 3 involves a comparison of models to data (Figure 3c). (iii) Cases studies 4 and 5 use manual refinement to fit experimental data to a fuzzy (case 4) or Boolean (case 5) logic-based model (Figure 3d). (iv) Case study 6 presents a formal method for model optimization

based on refining a literature-based Boolean model against high-throughput data (Figure 3e).

Case Study 1: Boolean Logic Model of Leukemic T Cell Large Granular Lymphocytes (29). Zhang et al. use a Boolean network model constructed from the literature to ask which proteins in leukemic T cell large granular lymphocytes (T-LGL) should be inhibited to induce apoptosis. Simulation of a 58-node logic model of the T-LGL survival signaling network is used to address the following questions. (i) What are minimal stimulation conditions that recapitulate observed deregulation of the T-LGL network? (ii) What perturbations might reverse deregulation and promote apoptosis?

A literature survey and experimental observations were combined to assemble a Boolean logic network describing signaling in T-LGL that affected cytoskeleton signaling, apoptosis, and proliferation. Simulations were compared when all nodes were free to vary and when some nodes were fixed (i.e., set to active or inactive and not allowed to change during the asynchronous updates). When the appropriate nodes were fixed, the model correctly recapitulated the situation in which leukemic T-LGL failed to undergo activation-induced cell death. Model analysis predicted a minimum set of stimuli that would result in the deregulated survival signaling previously observed in leukemic T-LGL. Experimental inhibition of this network state was shown to induce apoptosis in leukemic but not normal peripheral blood mononuclear cells (PBMC). Intriguingly, the authors identified nodes whose activation or inactivation caused the apoptosis node to be activated. These nodes are potential therapeutic targets for induction of apoptosis in leukemic T-LGL. Chemical knockdown
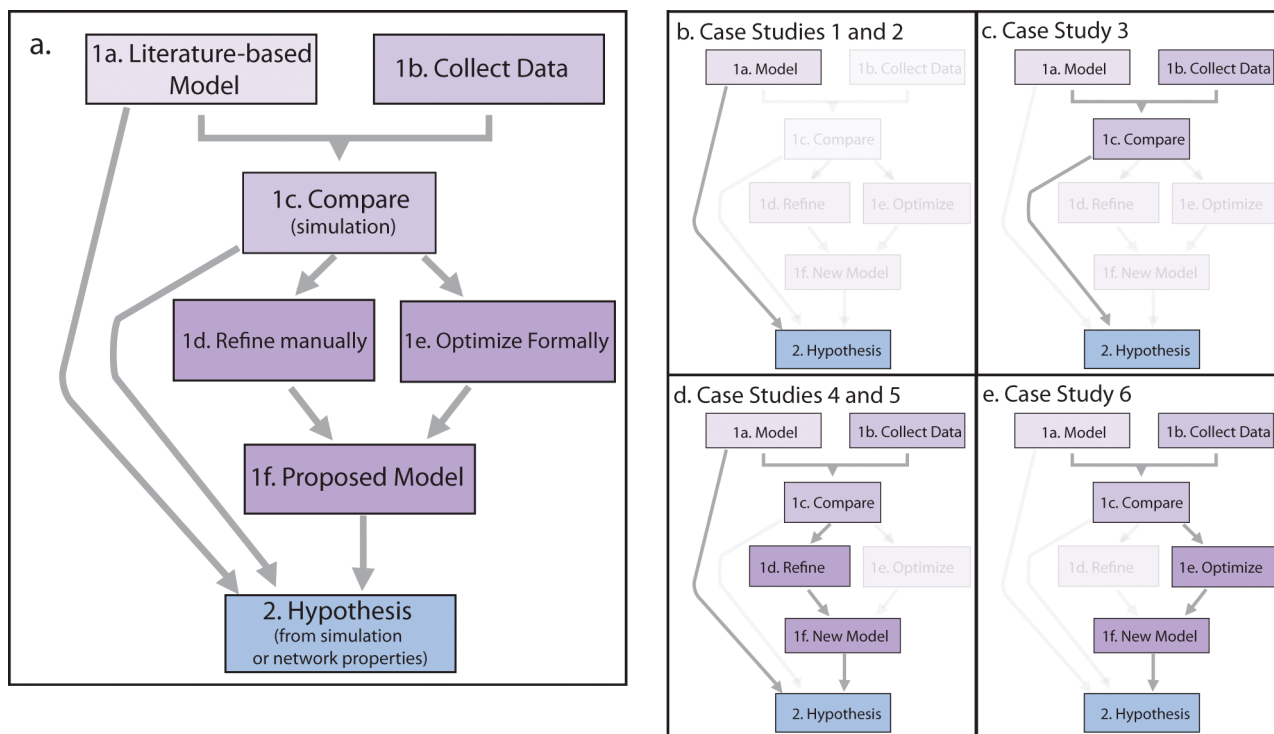
FIGURE 3: Workflow of application of logic-based models to answer biological questions. (a) General workflow. The workflow is divided into two phases: an initial model building phase (purple boxes) and a model prediction phase (blue box). Hypotheses are made from models built either from literature (box 1a) or from a comparison of a literature-based model with data (boxes 1a−c). In some cases, the models are refined manually (box 1d) or optimized formally (box 1f) with data and then used to make hypotheses (boxes 1a−f). (b) Workflow of case studies 1 and 2. These case studies analyze network properties of logic-based models built from the literature and use them to make experimentally testable predictions. (c) Workflow of case study 3. This case study compares the results of experiments to simulation results of a logic-based network to make predictions. They also analyze the network properties of their logic-based network. (d) Workflow of case studies 4 and 5. Both case studies manually refine their models based on experimental data, and prior to refinement, case study 5 first uses a model built from the literature to predict experimental outcome. (e) Workflow of case study 6. This case study compares logic-based models to experimental data and presents a formal method of training a Boolean network model to data.

of two of the identified nodes, sphingosine kinase 1 and NFκB, did indeed result in an increased level of apoptosis in leukemic T-LGL but not normal PBMC.

Case Study 2: Logic-Based Model of Helper T Cell Differentiation (22). Mendoza (22) used a literature-derived logic network model of interactions among five cytokines and transcription factors in helper T cells (Th cells) to ask the following questions. (i) Do the final states of a logic-based network correctly represent the differentiation fates of the helper T cell (Th cell)? (ii) How do feedback loops in cytokine signaling interact to generate specific cell fates? (iii) How do perturbing nodes of the logic network change the differentiation fate of Th cells?

A 17-node logic-based model of the Th regulatory network was constructed from published literature and simulated under all combinations of initial node states until logic steady states were achieved. This analysis revealed four steady states: one corresponding to Th0 cells, one corresponding to Th2 cells, and two corresponding to Th1 cells. The Th1 cell attractors differed in their level of secretion of IFNγ, but their level of IFNγ receptor was the same, a result supported by the literature. The feedback circuits that caused the network to reach these steady states were shown to correspond to experimental conditions known to induce Th0 cells to differentiate into Th1 or Th2 cells. Moreover, literature data validated several predictions based on single-node perturbations that corresponded to deletion or overexpression.

This paper illustrates the utility of logic-based modeling when analyzing a network involving many positive and negative

interactions whose net effect is not intuitively obvious. This type of model could be used to answer a number of interesting biological questions. For example, after a cell has entered one steady state, what cytokines or inhibitors must be present to force it to undergo a switch to another state? How might systemic cytokine administration affect the Th cell population? Can manipulation of normal nodes compensate for defects in nodes mutated in disease?

Case Study 3: Boolean Logic Model of ErbB Receptor Phosphoprotein Signaling Data (27). In the examples cited above, no direct link exists between the construction of the logic-based model and experimental data (Figure 3b). In contrast, Samaga et al. directly compared the outputs of a Boolean logic model constructed from the literature to data collected from cells (Figure 3c). The authors first developed a strategy for converting a biochemical network into Boolean logic. They then used this method to construct a complex Boolean logic model from a canonical graph of ErbB signaling that has been assembled by Kitano and colleagues (48). Finally, they asked the following question: Is the constructed Boolean model consistent with data from cells stimulated with ErbB ligands?

Model construction and simulation by Samaga et al. (27) were performed using the toolboxes ProMoT (49) and CellNetAnalyzer (CNA) (50), and data were obtained by exposure of HepG2 liver cancer cells and primary human hepatocyte to various ErbB ligands in the presence and absence of specific small molecule kinase inhibitors. Inconsistencies between model prediction and experimental observation generated a set of 11 hypotheses

regarding ErbB signaling in HepG2 and primary cells. Five of the 11 were supported by literature (although not in the cell types used in this study); five pointed to the need to remove or add interactions in the network, and one suggested that a small molecule inhibitor did not have the expected specificity. Significantly, this work successfully converted a biochemical map into an executable logic-based model and then used experiments to explore model topology.

Case Study 4: Fuzzy Logic Model of Protein Signaling Data (14). As a means of analyzing a set of continuous data, Aldridge et al. (*14*) built a fuzzy logic model of multiple growth factor and cytokine pathways based on prior literature knowledge and then refined the model manually on the basis of measurements of signaling protein phosphorylation in cells treated with TNFα, EGF, and insulin. During the model building process, the authors asked the following question: What interactions between TNFα and growth factors best explain the experimental data?

These data consisted of total or phosphoprotein levels for 11 signaling proteins following exposure of cells to TNFα, EGF, and insulin individually or in combination at 13 time points from 0 to 24 h. Because Boolean logic was unable to capture important intermediate states of protein modification in the data, fuzzy logic modeling was used. Fully implemented fuzzy logic is much more flexible than Boolean logic. Thus, the authors first selected a limited number of ways to represent interactions. Manual data fitting was used to optimize the interactions in the model and the shapes of the functions relating input and output species in the fuzzy logic gates. Time was included as a variable ("early" or "late"), and time delays were included in the logical rules for several gates. Acceptable values for these delays were determined manually. During the model building process, the authors uncovered unexpected interactions between ERK and IKK activities. This work demonstrates that fuzzy logic can be used to model and gain insight into signaling data that was not obvious from either inspection or partial least-squares regression modeling. The authors also note that because manual fitting of large data sets to a fuzzy logic model is an arduous process, methods are required to automate the fitting process.

Case Study 5: Integration of Logic-Based Modeling with Experimental Study of Trastuzumab-Resistant Breast Cancer (26). Sahin et al. first used a literature-derived Boolean logic model of a chemotherapeutic resistant cell line to ask the following question: Knockdown of what molecular species will result in increased drug sensitivity? Because the model was unable to accurately predict experimental results, they attempted to deduce the network from experimental data but concluded that the most reliable network was one that they had manually refined (Figure 3c).

Trastuzumab is a monoclonal antibody against ErbB2 that has successfully treated a subset of ErbB2 positive breast cancers. However, two-thirds of patients are Trastuzumab-resistant from the beginning of treatment. The authors hypothesize that this resistance is conferred by an escape from G1 cell cycle arrest. A Boolean logic network model of ErbB receptor regulation of the G1−S cell cycle transition was constructed on the basis of published literature. Only the ErbB receptor dimerization events that were possible in the cell line model of Trastuzumab resistance were included in the model, and initial node states were set on the basis of the biological activity of the proteins in their experimental system, making the model specific to the experimental system of interest, a clear benefit for modeling a context-sensitive phenomenon such as Trastuzumab resistance, which is context sensitive.

The retinoblastoma protein (Rb) is phosphorylated under conditions of constant EGF stimulus and was postulated to allow cells to escape G1 cell cycle arrest. The model was queried to identify those nodes whose inactivation under conditions of constant EGF would result in pRb dephosphorylation and consequent G1 cell cycle arrest (resulting in restoration of Trastuzumab sensitivity). RNAi knockdown of all but two species in the network (including those not predicted a priori to confer Trastuzumab sensitivity) was then used to test model-based predictions, several of which were found to be correct. Manually refining a single logical rule substantially improved accuracy, correctly predicting all but one RNAi knockdown result. The authors attempted to reverse engineer the network using protein array data but were unable to explain this final inconsistency. Overall, this work nicely illustrates the power of integrating experimental and logic-based modeling to gain a more complete understanding of the system of interest. As with case study 4, it also points to a need for more reliable methods of training of logic-based networks.

Case Study 6: Training a Boolean Logic Model of HepG2 Signaling (25). The primary advance of Saez-Rodriguez et al. (*25*) is the development of a formal method for optimizing logic-based models against experimental data, implemented in *CellNetOptimizer*. The data in this case were fairly extensive, comprising ∼1000 phosphoprotein measurements of 16 signaling proteins in tumor cells stimulated with one of six growth factors or inflammatory cytokines (TGFα, IGF1, TNFα, IL1α, LPS, and IL6) in the presence or absence of one of seven small molecule kinase inhibitors. The starting point for model construction was a signed directed graph comprising 82 nodes and 116 interactions derived from pathways in Ingenuity IPA. The authors then asked the following questions: (i) Can a formal training process be developed to increase the predictive capacity of the naïve model? (ii) Is the number of interactions in the optimized network similar to or smaller than the number in the naïve model? (iii) Can interactions that increase predictive power although they were absent from the initial graph be identified? It was observed that data-optimized models contain many fewer interactions than the original network graph, suggesting the presence of many false-positive interactions at least for the HepG2 cells under study. Moreover, addition of a small number of links deduced directly from data improved predictive capacity while increasing model size only modestly. Support for these links was subsequently found in the literature. This work represents a first step in using logic-based models to generate executable models of network graphs and then refining the models to increase their reliability in specific cellular contexts. Direct extension of the methods should make it possible to compare different cell types directly and perhaps even identify drugs that affect diseased but not normal cells.

CONCLUSIONS

In this work, we describe how logic-based models can be used to represent biochemical signaling networks and illustrate some of the questions that logic-based modeling can address. The ability of discrete and fuzzy logic models to determine the effects of protein overexpression or inhibition on phenotype, elucidate network properties, and identify the network that best describes high-throughput experimental data has been illustrated with case studies. We expect continued development to further the utility of logic-based modeling while also pointing to limitations in the

types of questions one can address with these models. For example, the ability to convert discrete logic-based models into continuous forms provides a means of investigating the dynamics of networks in which levels vary in a graded manner. A significant opportunity exists to determine the most effective way to apply the power of logic-based models to different biological networks of interest and answer basic questions. (i) What modeling formalism (e.g., logic-based, ODE, or regression-based) reaches the correct balance between too much and too little detail for each biological systems of interest? (ii) What modeling formalism yields the most interpretable results for each biological system of interest? (iii) Can logic-based models be embedded in more complex models to create a hybrids that represent some reactions in great detail and others in a more abstract manner? The recently demonstrated ability to train logic-based models on experimental data will also make it possible to tailor logic-based models to specific cell types and conditions, thereby providing a framework for predicting the effect of pharmaceuticals, mutations, and cell microenvironment on cell state.

ACKNOWLEDGMENT

SUPPORTING INFORMATION AVAILABLE

Overview of quantitative logic-based descriptions (Supplementary Figure 1), Matlab simulation code for binary simulation of an example network (Figure 1) (binaryEx_synch.txt), Matlab simulation code for binary simulation of an example network using truth tables (booleanEx_synch.txt), Matlab simulation code for multistate discrete synchronous simulation of an example network (multiEx_synch.txt), and Matlab simulation code for multistate discrete asynchronous simulation of an example network (multiEx_asynch.txt). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Kestler, H. A., Wawra, C., Kracher, B., and Kuhl, M. (2008) Network modeling of signal transduction: Establishing the global view. *BioEssays 30*, 1110–1125.
2. Heinrichs, A., Kritikou, E., Pulverer, B., and Raftopoulou, M., Eds. (2006) Systems biology: A user's guide, Nature Publishing Group, New York.
3. Janes, K. A., and Lauffenburger, D. A. (2006) A biological approach to computational models of proteomic networks. *Curr. Opin. Chem. Biol. 10*, 73–80.
4. Ideker, T., and Lauffenburger, D. (2003) Building with a scaffold: Emerging strategies for high- to low-level cellular modeling. *Trends Biotechnol. 21*, 255–262.
5. Aldridge, B. B., Burke, J. M., Lauffenburger, D. A., and Sorger, P. K. (2006) Physicochemical modelling of cell signalling pathways. *Nat. Cell Biol. 8*, 1195–1203.
6. Raj, A., and van Oudenaarden, A. (2008) Nature, Nurture, or Chance: Stochastic Gene Expression and Its Consequences. *Cell 135*, 216–226.
7. Bauer-Mehren, A., Furlong, L. I., and Sanz, F. (2009) Pathway databases and tools for their exploitation: Benefits, current limitations and challenges. *Mol. Syst. Biol. 5*, 290.
8. Rual, J., Venkatesan, K., Hao, T., Hirozane-Kishikawa, T., Dricot, A., Li, N., Berriz, G. F., Gibbons, F. D., Dreze, M., Ayivi-Guedehoussou, N., Klitgord, N., Simon, C., Boxem, M., Milstein, S., Rosenberg, J., Goldberg, D. S., Zhang, L. V., Wong, S. L., Franklin, G., Li, S., Albala, J. S., Lim, J., Fraughton, C., Llamosas, E., Cevik, S., Bex, C., Lamesch, P., Sikorski, R. S., Vandenhaute, J., Zoghbi, H. Y., Smolyar, A., Bosak, S., Sequerra, R., Doucette-Stamm, L., Cusick, M., Hill, D., Roth, F. P., and Vidal, M. (2005) Towards a proteome-scale map of the human protein-protein interaction network. *Nature 437*, 1173–1178.
9. Joughin, B. A., Cheung, E., Karuturi, R., Saez-Rodriguez, J., Lauffenburger, D. A., and Liu, E. T. (2009) Cellular Signaling Networks. In Systems Biomedicine, Concepts and Perspective (Liu, E., and Lauffenburger, D. A., Eds.) Elseveier, Amsterdam.
10. Kauffman, S. A. (1969) Metabolic stability and epigenesis in randomly constructed genetic nets. *J. Theor. Biol. 22*, 437–467.
11. de Jong, H. (2002) Modeling and simulation of genetic regulatory systems: A literature review. *J. Comput. Biol. 9*, 67–103.
12. Thomas, R., and D'Ari, R. (1990) Biological Feedback, CRC Press, Boca Raton, FL.
13. Huang, S., and Ingber, D. E. (2000) Shape-dependent control of cell growth, differentiation, and apoptosis: Switching between attractors in cell regulatory networks. *Exp. Cell Res. 261*, 91–103.
14. Aldridge, B., Saez-Rodriguez, J., Muhlich, J., Sorger, P., and Lauffenburger, D. A. (2009) Fuzzy logic analysis of kinase pathway crosstalk inTNF/EGF/insulin-induced signaling. *PLoS Comput. Biol. 5*, e1000340.
15. Franke, R., Muller, M., Wundrack, N., Gilles, E. D., Klamt, S., Kahne, T., and Naumann, M. (2008) Host-pathogen systems biology: Logical modeling of hepatocyte growth factor and *Helicobacter pylori* induced c-Met signal transduction. *BMC Syst. Biol. 2*, 4.
16. Gupta, S., Bisht, S. S., Kukreti, R., and Jain, S. (2007) Boolean network analysis of a neurotransmitter signaling pathway. *J. Theor. Biol. 244*, 463–469.
17. Helikar, T., Konvalina, J., Heidel, J., and Rogers, J. A. (2008) Emergent decision-making in biological signal transduction networks. *Proc. Natl. Acad. Sci. U.S.A. 105*, 1913–1918.
18. Huang, Z., and Hahn, J. (2009) Fuzzy modeling of signal transduction networks. *Chem. Eng. Sci. 64*, 2044–2056.
19. Kaufman, M., Andris, F., and Leo, O. (1999) A logical analysis of T cell activation and anergy. *Proc. Natl. Acad. Sci. U.S.A. 96*, 3894–3899.
20. Li, S., Assmann, S. M., and Albert, R. (2006) Predicting essential components of signal transduction networks: A dynamic model of guard cell abscisic acid signaling. *PLoS Biol. 4*, 1732–1748.
21. Mai, Z., and Liu, H. (2009) Boolean network-based analysis of the apoptosis network: Irreversible apoptosis and stable surviving. *J. Theor. Biol. 259*, 760–769.
22. Mendoza, L. (2006) A network model for the control and differentiation process in Th cells. *BioSystems 84*, 101–114.
23. Philippi, N., Wlter, D., Schlatter, R., Ferreira, K., Ederer, M., Sawodny, O., Timmer, J., Borner, C., and Dandekar, T. (2009) Modeling system states in liver cells: Survival, apoptosis, and their modifications in response to viral infection. *BMC Syst. Biol. 3*, 97.
24. Saez-Rodriguez, J., Simeoni, L., Lindquist, J. A., Hemenway, R., Bommhardt, U., Arndt, B., Haus, U. U., Weismantel, R., Gilles, E. D., Klamt, S., and Schraven, B. (2007) A logical model provides insights into T cell receptor signaling. *PLoS Comput. Biol. 3*, e163.
25. Saez-Rodriguez, J., Alexopoulos, L. G., Epperlein, J., Samaga, R., Lauffenburger, D. A., Klamt, S., and Sorger, P. (2009) Discrete logic modeling as a means to link protein signaling networks with functional analysis of mammalian signal transduction. *Mol. Syst. Biol. 5*, 331.
26. Sahin, O., Frohlich, H., Lobke, C., Korf, U., Burmester, S., Majety, M., Mattern, J., Schupp, I., Chaouiya, C., Thieffry, D., Poustka, A., Wiemann, S., Beissbarth, T., and Arlt, D. (2009) Modeling ERBB receptor-regulated G1/S transition to find novel targets for de novo trastuzumab resistance. *BMC Syst. Biol. 3*, 1.
27. Samaga, R., Saez-Rodriguez, J., Alexopoulos, L. G., Sorger, P. K., and Klamt, S. (2009) The Logic of EGFR/ErbB Signaling: Theoretical Properties and Analysis of High-Throughput Data. *PLoS Comput. Biol. 5*, e1000438.
28. Wittmann, D., Krumsiek, J., Saez-Rodriguez, J., Lauffenburger, D. A., Klamt, S., and Theis, F. (2009) From Qualitative to Quantitative Modeling. *BMC Syst. Biol. 3*, 98.
29. Zhang, R., Shah, M., Yang, J., Nyland, S., Liu, X., Yun, J., Albert, R., and Loughran, T. (2008) Network model of survival signaling in large granular lymphocyte leukemia. *Proc. Natl. Acad. Sci. U.S.A. 105*, 16308.
30. Zielinski, R., Przytycki, P. F., Zheng, J., Zhang, D., Przytycka, T. M., and Capala, J. (2009) The crosstalk between EGF, IGF, and insulin cell signaling pathways: Computations and experimental analysis. *BMC Syst. Biol. 3*, 88.
31. Wu, M., Yang, X., and Chan, C. (2009) A Dynamic Analysis of IRS-PKR Signaling in Liver Cells: A Discrete Modeling Approach. *PLoS One 4*, No. e8040.
32. Schlatter, R., Schmich, K., Vizcarra, I. A., Scheurich, P., Sauter, T., Borner, C., Ederer, M., Merfort, I., and Sawodny, O. (2009) ON/OFF

and Beyond: A Boolean Model of Apoptosis. *PLoS Comput. Biol. 5*, No. e1000595.

33. Mitsos, A., Melas, I., Siminelaki, P., Chairakai, A., Saez-Rodriguez, J., and Alexopoulos, L. G. (2009) Identifying Drug Effects via Pathway Alterations usingan Interger Linear Programming Optimization Formulation on Phosphoproteomic Data. *PLoS Comput. Biol. 5*, No. e1000591.

34. Thomas, R., and Kaufman, M. (2001) Multistationarity, the basis of cell differentiation and memory. I. Structural conditions of multi-stationarity and other nontrivial behavior. *Chaos 11*, 170–179.

35. Kaufman, M., Urbain, R., and Thomas, R. (1985) Towards a logical analysis of the immune response. *J. Theor. Biol. 114*, 527–561.

36. Thakar, J., Pilione, M., Kirimanjeswara, G., Harvill, T. T., and Albert, R. (2007) Modeling systems-level regulation of host immune responses. *PLoS Comput. Biol. 3*, e109.

37. Glass, L., and Kauffman, S. A. (1973) The Logical Analysis of Continuous, Non-linear Biochemical Control Networks. *J. Theor. Biol. 39*, 103–129.

38. Mendoza, L., and Xenarios, I. (2006) A method for the generation of standardized qualitative dynamical systems of regulatory networks. *Theor. Biol. Med. Modell. 3*, 13.

39. Thomas, R. (1973) Boolean formalization of genetic control circuits. *J. Theor. Biol. 42*, 563–585.

40. Garg, A., Di Cara, A., Xenarios, I., Mendoza, L., and De Micheli, G. (2008) Synchronous versus asynchronous modeling of gene regulatory networks. *Bioinformatics 24*, 1917–1925.

41. Klamt, S., Saez-Rodriguez, J., Lindquist, J., Simeoni, L., and Gilles, E. D. (2006) A Methodology for the Structural and Functional Analysis of Signaling and Regulatory Networks. *BMC Bioinf. 7*, 56.

42. Chaves, M., Sontag, E. D., and Albert, R. (2006) Methods of robustness analysis for Boolean models of gene control networks. *IEEE Proc. Syst. Biol. 153*, 154–167.

43. Fauré, A., Naldi, A., Chaouiya, C., and Thieffry, D. (2006) Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics 22*, e124–e131.

44. Shmulevich, I., Dougherty, E. R., Kim, S., and Zhang, W. (2002) Probabilistic Boolean Networks: A rule-based uncertainty model for gene regulatory networks. *Bioinformatics 18*, 261–274.

45. Gat-Viks, I., and Shamir, R. (2007) Refinement and expansion of signaling pathways: The osmotic response network in yeast. *Genome Res. 17*, 358–367.

46. Fisher, J., and Henzinger, T. (2007) Executable cell biology. *Nat. Biotechnol. 25*, 1239–1249.

47. Abdi, A., Tahoori, M. B., and Emamian, E. S. (2008) Fault diagnosis Engineering of Digital Circuits Can Identify Vulnerable Molecules in Complex Cellular Pathways. *Sci. Signal. 1*, ra10.

48. Oda, K., Matsuoka, Y., Funahashi, A., and Kitano, H. (2005) A comprehensive pathway map of epidermal growth factor receptor signaling. *Mol. Syst. Biol. 1*, No. 2005 0010.

49. Saez-Rodriguez, J., Mirschel, S., Hemenway, R., Klamt, S., Gilles, E. D., and Ginkel, M. (2006) Visual setup of logical models of signaling and regulatory networks with ProMoT. *BMC Bioinf. 7*, 506.

50. Klamt, S., Saez-Rodriguez, J., and Gilles, E. D. (2007) Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst. Biol. 1*, 2.

51. Li, F., Long, T., Lu, Y., Ouyang, Q., and Tang, C. (2004) The yeast cell-cycle network is robustly designed. *Proc. Natl. Acad. Sci. U.S.A. 101*, 4781–4786.

52. Davidich, M. I., and Bornholdt, S. (2008) Boolean network model predicts cell cycle sequence of fission yeast. *PLoS One 3*, e1672.

53. Mendoza, L., and Alvarez-Buylla, E. R. (1998) Dynamics of the genetic regulatory netowrk for *Arabidopsis thaliana* flower morphogenesis. *J. Theor. Biol. 193*, 307–319.

54. Mendoza, L., Thieffry, D., and Alvarez-Buylla, E. R. (1999) Genetic control of flower morphogenesis in *Arabidopsis thaliana*: A logical analysis. *Bioinformatics 15*, 593–606.

55. Braunewell, S., and Bornholdt, S. (2007) Superstability of the yeast cell-cycle dynamics: Ensuring causality in the presence of biochemical stochasticity. *J. Theor. Biol. 245*, 638–643.

56. Albert, I., Thakar, J., Li, S., Zhang, R., and Albert, R. (2008) Boolean network simulations for life scientists. *Source Code Biol. Med. 3*, 16.

57. Chaouiya, C., Remy, E., Mosse, B., and Thieffry, D. (2003) Qualitative Analysis of Regulatory Graphs: A Computations Tool Based On a Discrete Formal Framework. In Lecture Notes in Control and Information Sciences, Positive Systems (Benvenuit, L., De Santis, A., and Farina, L., Eds.) pp 119−126, Springer-Verlag, Berlin.

58. Gonzalez, A. G., Naldi, A., Sanchez, L., Thieffry, D., and Chaouiya, C. (2006) GINsim: A software suite for the qualitative modelling, simulation, and analysis of regulatory networks. *BioSystems 84*, 91–100.

59. de Jong, H., Geiselmann, J., Hernandez, C., and Page, M. (2003) Genetic Network Analyzer: Qualitative simulation of genetic regulatory networks. *Bioinformatics 19*, 336–344.

60. Helikar, T., and Rogers, J. A. (2009) ChemChains: A platform for simulation and analysis of biochemical networks aimed to laboratory scientists. *BMC Syst. Biol. 3*, 58.

61. Ulitsky, I., Gat-Viks, I., and Shamir, R. (2008) MetaReg: A platform for modeling, analysis, and visualization of biological systems using large-scale experimental data. *Genome Biol. 9*, R1.

62. DiCara, A., Garg, A., DeMicheli, B., Xenarios, I., and Mendoza, L. (2007) Dynamic simulation of regulatory networks using SQUAD. *BMC Bioinf. 8*, 462.