Overview

Boolean modeling: a logic-based dynamic approach for understanding signaling and regulatory networks and for making useful predictions

Réka Albert[1*] and Juilee Thakar[2]

The biomolecules inside or near cells form a complex interacting system. Cellular phenotypes and behaviors arise from the totality of interactions among the components of this system. A fruitful way of modeling interacting biomolecular systems is by network-based dynamic models that characterize each component by a state variable, and describe the change in the state variables due to the interactions in the system. Dynamic models can capture the stable state patterns of this interacting system and can connect them to different cell fates or behaviors. A Boolean or logic model characterizes each biomolecule by a binary state variable that relates the abundance of that molecule to a threshold abundance necessary for downstream processes. The regulation of this state variable is described in a parameter free manner, making Boolean modeling a practical choice for systems whose kinetic parameters have not been determined. Boolean models integrate the body of knowledge regarding the components and interactions of biomolecular systems, and capture the system's dynamic repertoire, for example the existence of multiple cell fates. These models were used for a variety of systems and led to important insights and predictions. Boolean models serve as an efficient exploratory model, a guide for follow-up experiments, and as a foundation for more quantitative models. © 2014 Wiley Periodicals, Inc.

How to cite this article:
WIREs Syst Biol Med 2014, 6:353–369. doi: 10.1002/wsbm.1273

INTRODUCTION

A key aim of Systems Biology is to elucidate the emergent properties and behaviors of biological systems. For example, Systems Biology aims to explain how cellular behaviors such as movement or proliferation result from the interactions among sub-cellular components such as proteins and small molecules. High-throughput and targeted experiments provide a rich source of interactions.[1] Experiments should be complemented by computational and modeling approaches to obtain an understanding from the data[2] and to generate testable hypotheses. For example, representation as a network of nodes connected pairwise by edges offers a coherent representation of a system of interacting biomolecules.[1,3–6] Going further, network-based dynamic models describe how the abundances of the biomolecules in the network vary in time due to the interactions they participate in.

Boolean dynamic models were introduced as a prototypical model of gene regulatory networks.[7,8] After assembling the components of a system and their regulatory interactions, a Boolean model describes

*Correspondence to: rza1@psu.edu

[1]Department of Physics, Department of Biology, Pennsylvania State University, University Park, PA, USA

[2]Department of Microbiology and Immunology, Department of Biostatistics and Computational Biology, University of Rochester Medical Center, Rochester, NY, USA

Conflict of interest: The authors have declared no conflicts of interest for this article.

 © 2014 Wiley Periodicals, Inc.

each node with a binary valued state variable. The 0 (OFF) state refers to a concentration or activity that is insufficient to initiate downstream processes and the 1 (ON) state represents a sufficient (above-threshold) concentration or activity. Boolean models describe the qualitative temporal behavior of the system. They can be used to elucidate how perturbations may alter normal behavior and thus lead to testable predictions which are especially valuable in poorly understood large-scale systems.[9] Boolean models have been successfully applied to gene regulatory networks,[10–12] signal transduction networks,[13,14] and networks at the physiological[15–17] or even population level.[18] They generated fruitful insights and predictions for unicellular organisms,[19–21] plants,[22–24] animals,[15,25] and humans, especially human signaling networks associated to diseases.[26–29] In this article, we describe the key steps of Boolean modeling of biological regulatory networks. For each step, we discuss the challenges, decisions and revisions that may need to be made, and illustrate them with examples from the recent literature.

DYNAMIC MODELING OF BIOMOLECULAR NETWORKS

When representing a system of interacting components (e.g., biomolecules) as a network, the components of the system become nodes (vertices), and the interactions, and relationships among the nodes become edges (links). Edges in the network are usually directed, indicating the orientation of mass transfer or information propagation, and can also be distinguished by a positive or negative sign to represent activation or inhibition. This network representation is the basis for structural analysis and dynamic modeling of the biological system. Structural analysis of the network involves the use of graph-theoretical measures, such as centrality measures and shortest paths.[3] Dynamic modeling approaches can be continuous or discrete according to the characterization of the molecular abundances by continuous or discrete variables. Continuous dynamic modeling[5,30,31] requires the knowledge of mechanistic details for each interaction and its parameterization with, e.g., rate constants. Since for most systems the values of many of these parameters are unknown and difficult to estimate, continuous modeling is only practical for systems with up to a few dozens of components. Discrete dynamic modeling such as Boolean network models,[32,33] multivalued logical models,[34,35] and Petri nets,[36,37] does not use kinetic parameters and is able to provide a qualitative dynamic description of the system. These approaches are practical for systems with hundreds of components and have been increasingly used in modeling biological networks.[14,28,38,39]

We illustrate the transition from continuous to Boolean modeling[40] for the case of a messenger RNA (mRNA), X, that has a single transcriptional activator, Y. A frequently used continuous model[41] assumes that the rate of mRNA transcription depends on the concentration of the transcription factor as a nonlinear Hill function, and that mRNA degradation is uncatalyzed. Consequently, the rate of change in the mRNA concentration follows $\frac{d[X]}{dt} = T\frac{[Y]^n}{[Y]^n+H^n} - \gamma\,[X]$, where the square brackets denote concentration, T is the maximal transcription rate and γ is the mRNA degradation rate. H is the transcription factor concentration at which the rate of mRNA synthesis is half of the maximal rate, and n is the Hill coefficient, giving the slope of the Hill function around the $[Y]=H$, $[X]=T/2$ point. For simplicity of illustration we will assume in the following that $T=\gamma=1$. The blue symbols in Figure 1 illustrate the time-course of $[X]$ for a Hill function with $H=0.5$ and $n=5$. If the Hill coefficient is high (which can happen due to cooperativity among transcription factors, for example), the synthesis term can be approximated by a Boolean step function, $B([Y])=0$ if $[Y]<H$, $B([Y])=1$ if $[Y]\geq H$ (see the green symbols in Figure 1). This intermediate step between continuous and Boolean modeling is in fact a modeling approach in its own right, called a piecewise linear or hybrid model.[42] In a piecewise linear model each node has two variables: a continuous variable akin to a normalized concentration, and a discrete variable akin to an activity. Thus our model

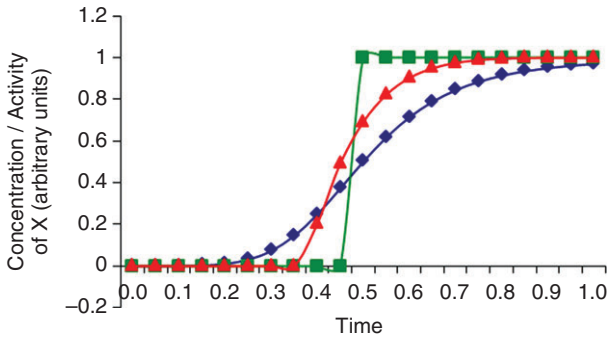


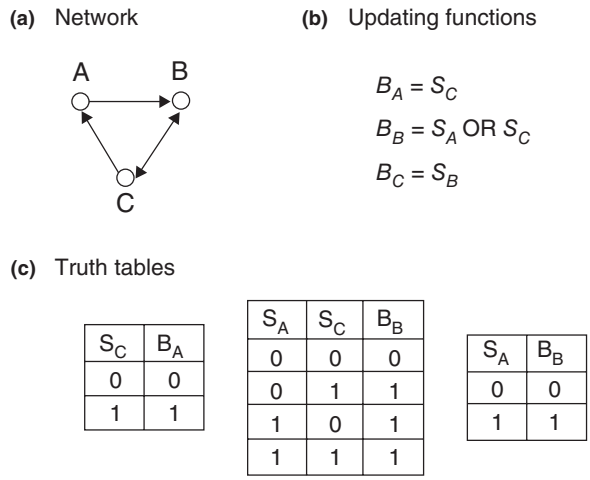
FIGURE 1 | Illustration of the transition between a continuous and a Boolean description of a process through which node Y positively regulates node X. It is assumed that the concentration of the node Y increases linearly in time, $[Y]=t$. The diagram indicates the time-course of the concentration $[X]$ or activity $\underline{X}$ of node X as a function of time at intervals of 0.05. Blue diamonds correspond to the case when the regulation is described by a Hill function, i.e., $\frac{d[X]}{dt} = \frac{(t)^n}{(t)^n+(0.5)^n} - [X]$, with $n=5$. Red triangles correspond to regulation described by a step function, i.e., $\frac{d[X]}{dt} = B(t) - [X]$, where $B(t)=1$ if $t\geq 0.5$ and is zero otherwise. Green squares correspond to a Boolean function $\underline{X}(t)=B(t)$.

 © 2014 Wiley Periodicals, Inc.

becomes $\frac{d[X]}{dt} = \underline{Y}\,[X]$, where $\underline{Y} = B([Y])$ is the discrete variable describing the activity of the transcription factor. When $\underline{Y} = 0$, $[X]$ will decay to zero (or stay at zero) and when $\underline{Y} = 1$, $[X]$ will increase asymptotically toward one (see the red symbols in Figure 1). The next degree of approximation is to describe the mRNA with a Boolean value as well. This Boolean value, $\underline{X}$, follows the value of $\underline{Y}$ with a time delay. Denoting the $\underline{X}$ value at a future time by $\underline{X}^*$, the Boolean relationship between the two nodes' states is $\underline{X}^* = \underline{Y}$.

Which type of model is most suitable depends on the quantitative detail of the available experimental data: continuous models are best when sufficient mechanistic and kinetic information is available, discrete models are most practical for poorly characterized systems with no kinetic information, and hybrid models can be used when partial mechanistic and kinetic information is available. The focus of this overview is Boolean modeling, which as we will see serves as a very effective first, exploratory model of biological systems.

Each node i of a Boolean network model stands for a sub-cellular component such as a gene, protein, ion channel, metabolite, or signaling molecule. In addition, biological outcomes such as apoptosis or stomatal closure can also be represented as nodes of the network. Each node is characterized by a binary state (representing expression level, concentration, or activity) S_i. $S_i = 1$ (ON) means that component i is expressed, has an above-threshold concentration, or is active, and $S_i = 0$ (OFF) denotes that it is not expressed, has a below-threshold concentration or is inactive. It is not necessary to know the thresholds invoked in the definition of states as long as it can be assumed that a concentration level exists above which the component in question can effectively regulate its downstream targets. In Boolean models, the future state of node i, S_i^*, is given by a logic statement using the current states of its regulators. This statement, called a Boolean update function (or Boolean rule), represents the conditional dependency among the input (regulator) nodes in regulating the downstream target node. This function is usually expressed via the logic operators AND, OR, and NOT. For example, B_D = (A OR B) AND NOT C is a Boolean function. An additional possibility is to use a threshold function, comparing a weighted sum of the inputs to a node-specific activation threshold.[18,20] A Boolean update function can also be represented by a truth table. The truth table of a Boolean function with k regulators has $k+1$ columns (one for each of the regulators, the last one for the target) and 2^k rows (one for each combination of state values of the k regulators). Figure 2 sketches a three-node network and the Boolean functions that express the regulation among the nodes.



S_C	B_A
0	0
1	1

S_A	S_C	B_B
0	0	0
0	1	1
1	0	1
1	1	1

S_A	B_B
0	0
1	1

FIGURE 2 | A simple Boolean network model. (a) The directed network associated with the Boolean model. The edges with arrows represent positive effects. Note that the network does not uniquely determine the Boolean updating function for node B. (b) The Boolean updating functions in the model. Note that the network can also support an alternative updating function, $B_B = S_A$ AND S_C. (c) The truth tables of the Boolean updating functions given in (b).

By successively re-evaluating each node's state by applying the corresponding Boolean update function, the system's collective state, i.e., $(S_1(t), S_2(t), \ldots, S_n(t))$, changes over time and ultimately reaches a fixed point (steady state) or a set of recurring states. This terminal state or set of states is called an attractor. The attractors of biomolecular interaction networks represent cellular phenotypes and behaviors.[20,43,44] The parameter free nature and qualitative features of Boolean modeling make it suitable for analyzing the repertoire of behaviors of a large-scale system, such as its possible multistability (the existence of multiple stable steady states),[45] the initial conditions that lead to one attractor versus the other, the activity changes of components following a perturbation,[46] and the stability of cellular responses to a signal.

THE MAIN STEPS OF CONSTRUCTING A BOOLEAN DYNAMIC MODEL OF A BIOMOLECULAR INTERACTION NETWORK

As illustrated in Figure 3, the model construction starts with a compilation of a list of components (nodes). Then, or at the same time, the interactions and regulatory relationships among these nodes need to be synthesized from the experimental literature. These interactions and relationships will become

 © 2014 Wiley Periodicals, Inc.

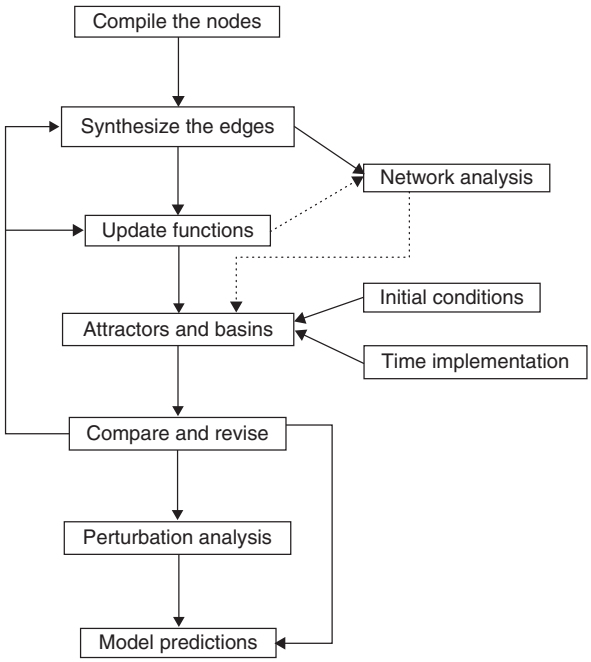


FIGURE 3 | Illustration of the main steps of constructing a Boolean dynamic model of a biomolecular interaction network. The directed edges among steps indicate the order in which they may be tackled. The dashed edges mark complementary analysis that is currently not routinely undertaken but we expect will be increasingly used.

the edges of the reconstructed interaction network. Graph-theoretical analysis of this network is a useful follow-up, because it can identify key nodes or interaction patterns. The model construction continues with determining the Boolean update functions for each node based on the edges incident on the node and also using information from the literature. A key step is to compress the biological knowledge regarding the concentration or activity of each node into two states, since the model will be using binary states. The collection of the resting or pre-stimulus states of the nodes will be used as an initial condition in the model. The information on the steady state or long-term dynamics of the system will be used for model validation. The model construction also includes a choice for the representation of time, which is an implicit variable in Boolean models. This choice has a subtle influence on certain outcomes of the model.

Having chosen the update functions, initial conditions and the representation of time, the model can be run and it will indicate how the system's state changes in time. This information can be summarized as a set of attractors, and the initial conditions that converge to each attractor (called the basin of the attractor). The model-indicated attractors need to be compared with the biological information on the steady states and long-term behaviors of the system. If there are qualitative differences that cast doubt on the model, the edges or update functions of the model need to be re-checked and suitably revised. Here it is worth stressing that the steady states or long-term behaviors of the system cannot be used as inputs to the model, only for model checking. The inputs to the model are embodied in the update functions of each node, which relate it to its closest upstream regulators. The same way as for the modeled system, the attractors of the model are emergent properties. Qualitative agreement between the model's results and biological knowledge gives confidence to the model and allows its use to attain a higher degree of understanding and to make new predictions. For example, an often used follow-up is a comprehensive analysis of the effects of node perturbations.

Also sketched in Figure 3, using dashed lines, is a methodology to integrate information from the network and from the Boolean update rules into a second network, whose analysis can indicate the model's attractors directly, without dynamic modeling. We will come back to this topic in the next section. In the following we present in detail the individual steps of the dynamic modeling process.

Compile Components

As currently it is unfeasible to comprehensively model the dynamics of genome-scale regulatory networks, most models focus on a single behavior or outcome, e.g., segmentation in *Drosophila melanogaster*,[41,47] or differentiation of T cells.[48,49] The model aims to include all the genes and gene products involved in the relevant outcome or behavior. The modeler usually starts from a core set known from the literature, and expands it by including additional information. For example, gene expression data can be used to identify which genes are differentially expressed over conditions relevant to the outcome of interest, thus indicating that these genes may be associated with the outcome.[50–52] Putative nodes of the regulatory network can be also identified from causal experiments, where manipulation (e.g., knockout) of an already known component of the network leads to variations in the expression of a gene or activity of a protein, implying that the gene or protein should be added to the network.[53]

Construct the Interaction Network

This step involves the translation of experimental information into edges, followed by assembly and refinement of these edges. The most useful information is physical or biochemical evidence indicating direct interaction between two components, and evidence

 © 2014 Wiley Periodicals, Inc.

of the causal effect of the genetic mutation or pharmacological inhibition of a component on another component.[54] The causal relationships can be represented as directed edges from one component to another, and can be characterized by one of two signs: activating (positive) or inhibitory (negative). Indirect relationships among the nodes of a regulatory network can be also inferred from analysis of gene expression, proteomic or metabolomic data, using probabilistic or deterministic methods.[55–61]

The integration of the indirect causal evidence is often challenging, because each such relationship may involve other, known or unknown nodes. In some cases, evidence from multiple experiments leads to multinode causal relationships which then need to be broken down to putative pairwise relationships.[54] Part of this process can be formalized and is implemented in the software package NET-SYNTHESIS,[62] which generates the sparsest network consistent with the given causal evidence. The best use of this software is in iteration with additional literature search until the most appropriate network representation of the available experimental observations is found.[26]

Analyse the Interaction Network

Structural analysis of the assembled network by means of graph-theoretical measures provides information on the importance of individual nodes/edges, characterizes network neighborhoods and sheds light on the global organization the network. The most often used measures include centrality measures (such as node degree or betweenness centrality) and connectivity measures (such as distance). Software packages for network visualization and analysis include yEd Graph Editor,[63] Cytoscape,[64,65] CellNetAnalyzer,[66,67] NetworkX,[68] and Pajek.[69]

Centrality measures describe the role of individual nodes in the network. For example, the node degree quantifies the number of edges connected to each node. In directed networks, the in-degree of a node is defined as the number of edges coming into the node and the out-degree is the number of edges going out of the node. In particular, the nodes with only outgoing edges (nodes with in-degree = 0) are called sources. These nodes act as initial points of the flow of mass or information in the network. The nodes with only incoming edges (out-degree = 0) are sinks of the network; they act as terminal points of flow. It is also useful to identify the nodes whose degree is highest among the nodes, termed hubs. These hub nodes, although rare, play an important role in the network. For example, their loss can break a network into isolated clusters.[3] Temporal expression patterns may be used to further classify hubs into permanent (also called party) hubs which interact with all of their partners at the same time, and transient (also called date) hubs which are influential in one condition but less so in others.[70,71] Permanent protein hubs have multiple binding sites, while transient hubs tend to have a single binding site.[72]

Connectivity measures are based on the concept of path, which is a sequence of adjacent edges in the network. An undirected network is connected if there is at least one path between every pair of nodes. A directed network is strongly connected if every pair of nodes (let's call them A and B) has two directed paths of opposite directions (one from A to B and one from B to A). For example, the network in Figure 2(a) is strongly connected. If a network is not (strongly) connected, one can search for (strongly) connected components (also called subgraphs) in the network. The absence of strongly connected components (SCCs) indicates an acyclic structure (i.e., the network does not contain feedback loops), while having a large SCC implies that the network has a central core. It was reported that the transcriptional regulatory networks have only small SCCs (e.g., three-node feedback loops),[73,74] whereas a large SCC was observed in metabolic[75] and signaling[76] networks. A strongly connected component may have an in-component (nodes that can reach the SCC) and out-component (nodes that can be reached from the SCC). Nodes in each of these subsets tend to have a common task. For example, in signaling networks, the nodes of the in-component tend to be involved in ligand-receptor binding and the nodes of the out-component are usually responsible for the transcription of target genes or for phenotypic changes.[76]

Network motifs, recurring patterns of interconnection with well-defined topologies,[77] are also informative. Frequently observed network motifs include feed-forward and feedback loops. For example, the network in Figure 2(a) contains both a three-node feed-forward loop (formed by the edges $C \rightarrow A$, $A \rightarrow B$, $C \rightarrow B$) and a three-node feedback loop (formed by the edges $A \rightarrow B$, $B > C$, $C \rightarrow A$). Feed-forward loops are more abundant in the transcriptional regulatory and signaling networks of different organisms than randomized networks that keep each node's degree.[77,78] Feed-forward loops support several functions such as filtering of noisy input signals, pulse generation, and response acceleration.[79] Feedback loops support multistability or oscillations.[33,80]

Determine the Boolean Functions

The next important step is to determine the Boolean functions of each node. This function will depend on

 © 2014 Wiley Periodicals, Inc.

the node's regulators indicated in the network. In addition, the function is informed by an interpretation of experimental observations in the literature regarding the node and its regulators, for example the expression of the regulated node when one of its regulators is knocked out. If a node has only one regulator, then a single variable appears in its Boolean rule. This variable is combined with a NOT operator if the regulator is an inhibitor. For example, consider a protein P whose activation (its ability to regulate downstream processes) requires its phosphorylation. We can represent the state of protein P as $S_P = 1$ (ON) if it is predominantly in the phosphorylated form, and as $S_P = 0$ (OFF) if it is predominantly in the unphosphorylated form. If P is solely regulated by a kinase K that phosphorylates it, the Boolean update function for the state of P is $B_P = S_K$, where S_K is the state of the kinase. If instead of a kinase the protein P solely interacts with and is dephosphorylated by a phosphatase R, the update function of P can be written as $B_P = \text{NOT } S_R$, indicating that in the absence of the phosphatase protein P is active, while in its presence it is inactive. An example of a node having a single regulator in Figure 2(a) is node A, which is activated by node C, resulting in the Boolean rule $B_A = S_C$.

The activation of many components requires two or more regulators. For example, the transcription process of a gene G may be activated by a transcriptional complex consisting of two proteins, P1 and P2. This can be represented by the AND operator describing the simultaneous presence of the two proteins: $B_G = S_{P1}$ AND S_{P2}. If a component, e.g., a protein with multiple phosphorylation sites, is positively regulated by multiple regulators and any of them can activate it independently, the independent effects of these regulators on the target component can be captured by the OR operator. For example, in the case of node B in Figure 2(a) the rule $B_B = S_A$ OR S_C expresses that either A or C can independently turn node B on.

For nodes with multiple regulators knowledge of the incoming edges (positive and negative regulators) does not uniquely determine the dependency relationships among node states. For example, node B in Figure 2(a) is regulated by A and C. How can one determine, in a real situation, if activation of both A and C, or only one of them, is required for the activation of B? If there is experimental evidence that knocking out either A or C leads to the absence of B, then AND should be used. Conversely, if there is evidence that only simultaneous knockout of A and C would inactivate B, then OR should be used. When such information is not available, one can construct several variants of the Boolean rules and determine the one that best reproduces the known properties of the real system.[81] If this is not feasible, the OR operator may be used as a default, and the model can be updated once additional information is obtained. Alternatively, one can employ probabilistic Boolean networks,[60,82] which incorporate uncertainty by assigning multiple Boolean rules to a node, each with a certain probability of being selected.

Determine the Initial Condition

Representation in a binary model requires that the biological information on the concentration or activity of the components (nodes) is compressed into two qualitative states. This may be done if we know the threshold concentration or activity at which the component is effective in regulating downstream processes. Then, below-threshold concentration or activity becomes the 0 state and above-threshold concentration or activity becomes the 1 state. We stress that the state designated as 0 does not mean the complete absence of that component. Since it is relatively rare to know such thresholds, comparison (e.g., by differential gene expression analysis) between two reference states (e.g., a healthy and a diseased state) can be the basis of the state assignment. If numerous reference states or state time-courses are available, one can cluster these states into two groups[83] and assign the state 0 to the lower-value group.

Ideally the model's starting state should be the biologically relevant resting or pre-stimulus state if it is known *a priori*. If the available information is insufficient, one can exhaustively explore all initial conditions wherein certain nodes are in a known state. As the total number of states of m unspecified-state nodes is 2^m, it may not be feasible to try all of these states, and sampling must be used instead. In either case a large number of replicate simulations should be done, and the results need to be summarized over these replicate simulations. For example, one calculates the fraction of realizations of a certain attractor. We can think of these replicate simulations as a population of cells which differ in their pre-stimulus states, and the fraction of realizations of an attractor can be interpreted as the probability that the system attains the corresponding cellular phenotype.[23,26]

Choose a Time Implementation

In most Boolean models, time is an implicit variable and the passing of time is implemented via synchronous or asynchronous update algorithms. Update means the determination of a node's next state based on its Boolean function and on the state of its regulators. In the simplest update scheme, called synchronous update, the states of all nodes are updated

 © 2014 Wiley Periodicals, Inc.

simultaneously according to the last state of the system.[32] Specifically, the state of node i at time step $t+1$, denoted by $S_i(t+1)$, is determined by the state of its k_i regulators at time t:

$S_i^{t+1} = B_i\left(S_{i_1}^t, S_{i_2}^t, \ldots, S_{i_{k_i}}^t\right)$, where B_i is the Boolean rule for node i and S_{i_j}, $1 \le j \le k_i$, are the states of its regulators. This type of update implicitly assumes that the timescales of all biological events in the system are similar and the state transitions of components are synchronized. Synchronous update leads to a deterministic state trajectory in which any system state can have at most one successor. However, the timescales of biological events can vary widely from fractions of seconds to hours,[1] and systems with a single timescale are rare. Asynchronous models aim to account for timing variation by updating the nodes in a nonsynchronous manner. There are deterministic asynchronous schemes with fixed individual timescales[10] or fixed time delays[84] and stochastic asynchronous schemes wherein each node is updated with a certain probability,[85] all nodes are updated according to a random sequence,[86] or one randomly selected node is updated at a time.[87] In stochastic asynchronous models, the same initial condition can lead to different successors due to the variability of the update scheme. Asynchronous schemes can be informed by existing knowledge about the relative timescales of components.[86,88,89] Updating schemes have a considerable effect on the dynamics of the system (see the next subsection). One can choose a scheme that is most realistic for the biological system of interest, or compare the results of different schemes on the same system.[44] In cases where there is no information to inform the choice of update scheme, updating one node at a time is the most effective choice.[44]

Attractor Analysis

The dynamic behavior of a Boolean model is determined by the Boolean regulatory functions of the nodes and is influenced by the chosen updating scheme. Starting from a chosen initial condition the system's state changes in time in discrete transitions, and finally stabilizes in an attractor (a limited set of states) representing the long-term behavior of the system. Attractors can be fixed points (steady states), wherein the state of the system does not change, or complex attractors (also called loose attractors), wherein the system's state keeps revisiting the same set of states. Each complex attractor of synchronous and deterministic asynchronous models is a repeated succession of states, called a limit cycle. The complex attractors of models that use stochastic asynchronous update or probabilistic Boolean functions may also be cyclic, but more often the states of the attractor are revisited irregularly. Fixed point attractors of regulatory and signaling networks correspond to the steady activation states of components associated with cellular phenotypes.[20,27,90] Complex attractors correspond to cyclic and oscillatory behaviors such as the cell cycle, circadian rhythms, or Ca^{2+} oscillations. Therefore, identifying the possible attractors is a biologically relevant and informative goal.

Trajectories starting in different initial conditions may lead the system to different attractors. Thus, it is very informative to construct a map of the possible trajectories of the system in state space. All possible states of the system, a total of 2^n, where n is the number of nodes, make up its state space. A compact representation of all possible trajectories is possible through the state transition graph, whose nodes are the states of the system and whose edges denote the allowed transitions among the states according to the chosen updating scheme. The state transition graph can be used to determine the attractors of the system, and to find the set of initial states that leads to a specific attractor, called the basin of attraction of that attractor. Fixed points will correspond to states that do not have any outgoing edges (transitions), only a loop (self-edge). Complex attractors form a terminal SCC of the state transition graph (i.e., an SCC with an empty out-component).[87]

Figure 4 represents the state transition graph of the network given in Figure 2 using synchronous update (a) and when updating one node at a time (b). In both models, states 000 and 111 are the fixed points (steady states). Indeed, as the fixed points of a system are time independent, they are the same for both synchronous and asynchronous update.[86] The choice of updating scheme can affect the probability with which the system reaches these fixed points when started from a given initial condition. In the synchronous model no other states converge into the state 000, while in the asynchronous model states 001, 100 and 010 can lead to the fixed point 000 in one step and states 101 and 110 can lead to 000 in two steps. Additionally, synchronous models may exhibit limit cycles which are not present in the corresponding asynchronous models.[44,88] By comparing the state transition graphs given in Figure 4(a) and (b), we see that in the state transition graph of the synchronous model, each state has a unique successor (i.e., it has an out-degree of one), which is not the case in the asynchronous model. Consequently, attractors of a synchronous model have disjoint basins of attraction, whereas the basin of attraction of different attractors in stochastic asynchronous models may overlap. For example, in Figure 4(b) five of the six states that are

 © 2014 Wiley Periodicals, Inc.

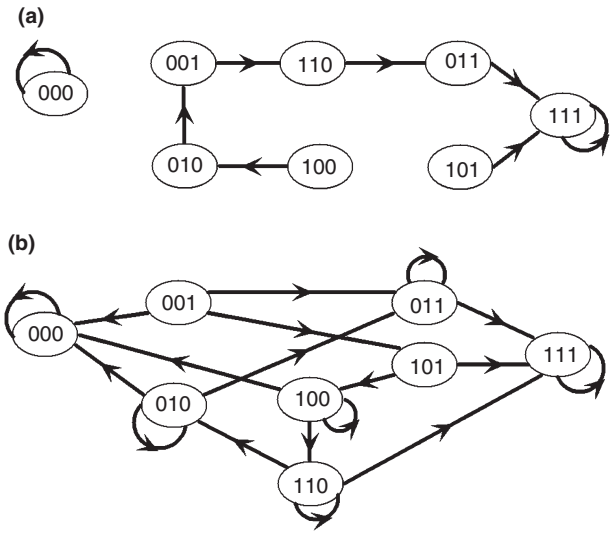


FIGURE 4 | State transition graphs corresponding to the Boolean model presented in Figure 2. The symbols correspond to the states of the system, indicated in the order A, B, C, thus 000 represents $S_A = 0$, $S_B = 0$, and $S_C = 0$. A directed edge between two states indicates the possibility of transition from the first state to the second by updating the nodes in the manner specified by the updating scheme. An edge that starts and ends at the same state (a loop) indicates that the state does not change during update. (a) The state transition graph corresponding to synchronous update, when all nodes are updated simultaneously. The two states that have loops are the fixed points of the system. (b) The state transition graph corresponding to updating one node at a time. While several states have loops, indicating that at least one of the nodes does not change state during its update, only the states that have no outgoing edges are fixed points of the system.

not fixed points are in the basins of attraction of both fixed points, and only state 011 is exclusive to the basin of 111.

An alternative method for determining the fixed points of small Boolean networks is to remove the time dependency from the Boolean rules and solve the resulting set of equations $B_i(S_1, \ldots, S_n) = S_i$ for all $1 \le i \le n$, where n is the number of nodes in the network. The solutions of this system of equations correspond to the fixed points of the Boolean model. For example, the fixed points of the simple network in Figure 2 can be obtained analytically by solving the following system of equations: $S_A = S_C$, $S_B = S_A$ OR S_C, $S_C = S_B$. Substituting the first equation into the second and simplifying the resulting equation using Boolean algebra results in $S_B = S_C$, which is equivalent with the third equation. Thus all three variables need to be equal, yielding two fixed points of the system, 000 and 111. Logical steady state analysis[67] can find the (partial) fixed points of Boolean models of signaling networks with sustained input signals by propagating the value of the input signals through the network.

Several software tools are available for Boolean dynamic modeling of biological systems. BooleanNet[91] can be used to simulate synchronous and stochastic asynchronous models and to determine the state transition graph. The R package BoolNet[92] provides attractor search and robustness analysis methods for synchronous, asynchronous and probabilistic Boolean models. SimBoolNet, a plugin to the widely used biological network analysis tool Cytoscape,[65] determines state trajectories and attractors using sequential update (starting from the external signals).

As a Boolean network model with n nodes has 2^n states, determining the attractor repertoire and state transition graph of a large system is a challenging problem. For larger networks, search methods that utilize the special features of attractors without the necessity of checking all possible trajectories have been developed.[90,93–95] An alternative solution is to use network reduction techniques to simplify the network while preserving its essential dynamical properties.[44,96–99] One method is to absorb a node by, simply said, connecting its upstream regulators to its downstream targets. Specifically, in the update function of all targets of the eliminated node one needs to replace the eliminated node's state by its update function. For example, in Figure 2(a) we can absorb node A, creating an edge from node C to node B. The updating rule of node B then becomes $B_B = S_C$ OR $S_C = S_C$. The network simplifies to a two-node positive feedback loop with still has two fixed points: 00 and 11. Indeed, iteratively absorbing nodes without a self-loop was proven to preserve the fixed points of a system[97,99] (as this is equivalent with eliminating variables while solving the set of equations $B_i(S_1, \ldots, S_n) = S_i$). However, this simplification may introduce spurious oscillations into the reduced model.[97] A more conservative network reduction method was shown to preserve both fixed points and complex attractors under stochastic asynchronous update.[100] This method is based on the propagation of the states of sustained signals and on the collapsing of nodes with in- and out-degree of one.[27,44]

Test the Correctness of the Model

The model must be able to reproduce prior experimental observations regarding input–output relations, dynamic behaviors, and cellular responses. For example, the model's attractor(s) need to match the biologically known steady state or oscillatory behavior of the system. For the purpose of this comparison the biological state of the system needs to be expressed in terms of binary variables. However, comparison with experimental observations is not limited

 © 2014 Wiley Periodicals, Inc.

to binary variables. When performing a large number of Boolean dynamic simulations the fraction of simulations that have a certain state is an intermediate value between 0 and 1. The shape of the time-course of this fraction can be qualitatively compared with the shape of experimental time-courses.[26]

If the model fails to reproduce known behaviors of the system, one needs to go back and check whether some important components or interactions are missing from the network structure, or whether some Boolean update functions are incomplete or wrong, e.g., use AND instead of OR or *vice versa*. The failure may also arise from the use of an inappropriate updating scheme (if the unmatched behavior is oscillatory) or initial condition (if a single initial condition was studied). A few rounds of revisions usually yield a Boolean dynamic model consistent with all known experimental observations. Many biological systems are robust,[20,41,47] so an indirect way to validate the model is to test its robustness to small perturbations such as interchanging OR and AND rules, switching the signs of interactions, scrambling interactions, adding or deleting a component or interaction. A good model can accommodate small perturbations, reflecting the adaptability of the system to diverse circumstances.

Generate Novel Predictions

While model construction and validation is time- and labour-intensive, the finished model is very valuable, as it transforms a set of separate facts into a system-level understanding. The power of Boolean dynamic modeling is its ability to predict the outcome repertoire of the system, generate testable hypotheses, and direct future wet-bench experiments in an efficient way. For example, the attractors of the system predict the activity of components in cellular responses or phenotype traits.[27,90,101] By analyzing the outcomes of the system from various initial conditions, we can understand how different signals (stimuli) crosstalk and lead to different cellular responses.

Once the attractors of the Boolean model of a system are determined, the activity of components in relevant cellular responses or phenotypes can be predicted. For example, the three fixed points of a Boolean model of a T-helper (Th) cell differentiation network[90] recapitulated the activation patterns of components observed in Th0, Th1, and Th2 cells, respectively. The attractors of the T cell activation induced cell death signaling network captured the normal (apoptosis) and disease (T-LGL) outcomes, and the latter identified the T-LGL state of components that were experimentally undocumented before.[27]

A Boolean model can be used to analyse the changes in the system's attractor repertoire in the case of system perturbations.[46,102] Knockout of a component can be simulated by fixing the corresponding node in the OFF state; constitutive expression can be simulated by fixing the node's state as ON. Transient perturbations can also be studied by implementing temporary (reversible) changes to the node's states. The model can predict the changes in the attractors of the system and their basins of attraction and identify the perturbations that lead to dramatic changes. For example, for the model in Figure 2, knockout of node A leads to the reduced system $S_A = 0$, $B_B = S_C$, $B_C = S_B$ which still has two attractors, the steady states 000 and 011. In contrast, knockout of node B leads to $B_A = S_C$, $S_B = 0$, $B_C = 0$, which only has one attractor, 000. Knockout of node C also leads to the 000 fixed point, while constitutively expressing any of the three nodes leads to a single attractor, the steady state 111. Thus only one of the permanent perturbations maintains the original system's capacity for having two attractors (bistability), the others destroy one of the two attractors. Perturbation analysis can identify the essential components that mediate phenotype traits.[15,23,26,27] For example, dynamic perturbation analysis for the T-LGL leukemia signaling network led to the identification of 19 potential therapeutic targets for the disease, more than half of which were supported by available experimental data or by follow-up experiments, and the rest can guide future experiments.[27]

We can also predict the biological role of regulatory interactions and feedback loops by removing them (by deleting a term from the update function of the target node) and comparing the dynamic sequences before and after the perturbations.[14,29,39] For example, disrupting any of five key feedback loops in a Boolean model of the p53 regulatory network led to the undesired outcome of cell death.[29] In summary, the model not only provides a systems-level understanding of the biological process, but also can direct follow-up targeted experiments.

INTEGRATED STRUCTURAL AND LOGICAL ANALYSIS OF BIOMOLECULAR NETWORKS

Boolean modeling allows the systematic identification of the dynamic repertoire of biological systems. However, for large systems that also lack timing information mapping of the state space is computationally expensive. Network reduction methods can help by reducing the size of the networks while preserving essential dynamic properties. Alternatively,

 © 2014 Wiley Periodicals, Inc.

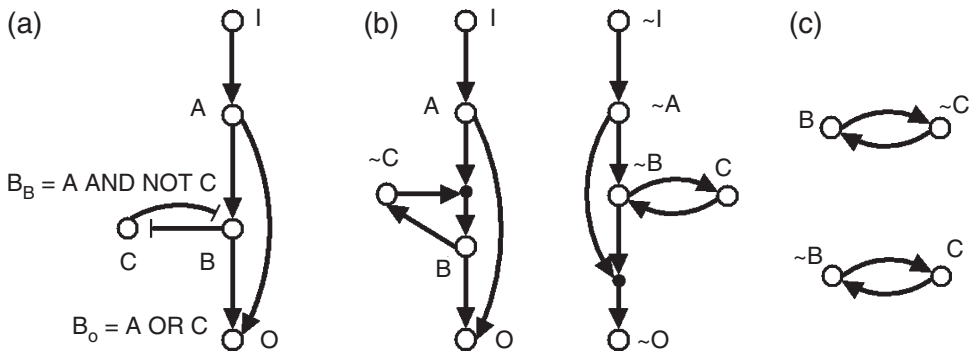


FIGURE 5 | Illustration of methods that integrate the structure and logic of regulatory interactions. (a) A hypothetical signal transduction network and Boolean model corresponding to it. The Boolean update rules that are not completely determined by the network are indicated next to the relevant nodes. (b) The expanded representation of the network, which includes three complementary nodes, indicated by prepending the node name by ~, and two composite nodes, indicated by small black filled circles. (c) The stable motifs of the expanded network in the case of a sustained input signal ($S_I = 1$). The first stable motif corresponds to the state 11101 (in the order I, A, B, C, O), while the second stable motif corresponds to the state 11011.

several methods have been developed to use the information encapsulated in the Boolean functions to enrich and improve structural analysis, potentially bypassing dynamical analysis.

One possibility, implemented in the software CellNetAnalyzer,[66] is to represent a biological network by a logical interaction hypergraph whose hyper-edges connect two sets of nodes instead of two nodes.[67,103] This way the relationship $B_C = S_A$ AND S_B can be represented by a hyper-edge that starts from the node set {A, B} and ends in C. An application of this hypergraph is in the identification of minimal intervention targets. A minimal intervention target is defined as a minimal set of nodes whose simultaneous manipulation satisfies a goal such as the permanent deactivation (off state) of the output node. These interaction targets are usually determined by systematic search, but analysis of the hypergraph can reduce the pool by eliminating candidate targets that cannot be successful and by grouping equivalent targets.

Our group developed an expanded network representation that integrates negation and conditional dependency among regulators into the network topology.[47,104] Specifically, the method introduces a complementary node for each node that is impacted by negative effects, and introduces a composite node for each set of interactions with conditional dependency. The new representation, wherein all interactions represent activation and all composite nodes indicate conditional dependency, facilitates a better functional interpretation of structural analysis. For example, the expanded network of the segment polarity genes provided important insights into the identification of coexpressed genes.[47] For example, it showed that the cells expressing *en* and *hh* never express *wg*, *ptc*, or *ci*, a well-known polarization that is the basis for the name 'segment polarity genes'.[105] Figure 5 illustrates the expanded network representation of a hypothetical signal transduction network composed of the input node I, intermediary nodes A, B, C and the output node O (Figure 5(a)). The Boolean update functions of the two nodes that have more than a single regulator are indicated. The expanded network (Figure 5(b)) is composed of two components that are disconnected from each other. The first component starts with the input node I ends in the output node O, and contains A, B, the complementary node of C, denoted as ~C, and a composite node shown as a black dot. The second component is made up by four complementary nodes, node C and a second composite node.

The new concept of elementary signaling mode was defined as the minimal set of components able to perform signal transduction independently. Figure 5(b) contains two elementary signaling modes that connect the input to the output: the path IAO, and the subgraph that contains the nodes I, A, ~C, the composite node (black dot) upstream of B, B, and O. We hypothesized that the signaling components whose disruption (and its cascading effects) eliminates the majority or all of the elementary signaling modes are essential. For example, in Figure 5(b) the loss of node A eliminates both elementary signaling modes between I and O, but the loss of node B leaves one of the elementary signaling modes intact. Validation on several signaling networks showed that this augmented structural method and essentiality criterion are in strong agreement with the results of dynamic simulations.[27,104]

The expanded network can also be used as a basis for network simplification. A structural criterion can be used to identify network motifs (subgraphs)

 © 2014 Wiley Periodicals, Inc.

that stabilize in a fixed state regardless of the rest of the network.[106] Specifically, a stable motif is the smallest strongly connected component in the expanded network which (1) does not contain both a node and its complementary node, and (2) if it contains a composite node, it also contains all of this node's input nodes. For example, in Figure 5(b) the nodes ~B and C form a stable motif. The fixed state of the nodes in the stable motif can be directly read out from the expanded network: if the stable motif contains a node, the node stabilizes in the ON state, and if the stable motif contains a complementary node, the corresponding node stabilizes in the OFF state. The fixed state of a network motif can be used for network simplification in a similar way as the sustained presence of a signal can. Iterative searching for stable motifs and network simplification leads to one of two possible outcomes: either there are no more nodes with unknown states, in which case a fixed point of the system is identified, or no new stable motifs are found, in which case the remaining nodes are expected to oscillate.

As an example of network simplification and attractor detection, let's consider the sustained presence of the input signal ($S_I = 1$) in Figure 5. The sustained signal leads to the stabilization of $S_A = 1$, and consequently of $S_O = 1$, while the regulation of B simplifies to $B_B = \text{not } S_C$ and similarly $B_C = \text{not } S_B$. The expanded representation of this mutual inhibition network consists of two disconnected positive feedback loops: one formed by B and ~C and the other by ~B and C (see Figure 5(c)). Both of these feedback loops are stable motifs, the first corresponding to $S_B = 1$ and $S_C = 0$ and the second to $S_B = 0$ and $S_C = 1$. Thus there are two fixed point attractors for this system which differ only in the state of nodes B and C.

Taken together, these integrated Boolean—structural studies revealed that while some properties of a dynamic model depend on initial conditions and individual timescales, other properties are encoded in the combinatorial regulations represented by Boolean rules and do not depend on the details of the dynamic simulation. Therefore, these integrated methods are fruitful as exploratory analysis of large networks where dynamic modeling is computationally impractical, or as a first step that guides follow-up targeted computational or experimental studies.

FROM BOOLEAN TO MORE QUANTITATIVE MODELING FRAMEWORKS

The use of Boolean modeling is most natural when the system's outcomes can be unambiguously categorized into two classes. Naturally, many applications exist where a binary outcome is a gross simplification and a continuous or multistate representation is needed. Indeed, there are multiple ways in which Boolean modeling can be improved. The construction of these improved models has similar steps as described earlier. The Boolean update functions are replaced by multistate or continuous functions, and in certain models time is continuous as well.

Several methods exist to connect between continuous and Boolean modeling. As we have seen before, piecewise linear models are a hybrid of Boolean and continuous models.[42] An example of a piecewise linear model is $\frac{d\hat{S}_i}{dt} = B_i\left(S_{i_1}, S_{i_2}, .., S_{i_{k_i}}\right) - \gamma_i \hat{S}_i$, $S_i = \begin{cases} 0, & \hat{S}_i < \theta_i \\ 1, & \hat{S}_i \geq \theta_i \end{cases}$, where $\hat{S}_i$ is the continuous variable (e.g., concentration) associated with node i, S_i is the discrete variable of node i, γ_i is a decay rate and θ_i is a threshold parameter. These types of models have been fruitfully applied due to their attractive combination of continuous time, quantitative information, and few kinetic parameters.[10,107,108] The parameters, such as activation thresholds, are at a higher, more coarse-grained level than the kinetics of elementary reactions. The software packages *Genetic Network Analyzer*[109] or BooleanNet[91] can be used to simulate piecewise linear models. Piecewise linear models retain the steady states of the corresponding Boolean model, and yield damped or sustained oscillations in cases where the Boolean model has a complex attractor.[42] Figure 6 illustrates the piecewise linear model of a negative feedback loop and compares it with a synchronous Boolean model in which all the nodes are updated at the same time. The synchronous Boolean model has two limit cycles, one containing six states and one containing two states (Figure 6(b)). In the piece wise linear model the decay rate of each node is assumed to be $\gamma = 1$ and the threshold is $\theta = 0.5$. Figure 6(c) indicates the time-courses of the continuous variables of the three nodes starting from the initial state $\hat{S}_A = \hat{S}_B = \hat{S}_C = 1$. The sustained oscillations exhibited by all three variables agree with the six-state Boolean limit cycle of Figure 6(b). Indeed, first $\hat{S}_A$ decays below the threshold (equivalent with the state 011), followed by $\hat{S}_B$, then $\hat{S}_C$ (equivalent with the state 000), then $\hat{S}_A$ comes back above the threshold (state 100), and so on.

The hybrid formalism called standardized qualitative dynamical system, implemented in the software SQUAD,[110] starts with a standardized Boolean update function for each node, which is then translated into a continuous sigmoidal function of weighted activating and inhibiting interactions. The steady states of the Boolean model are used as starting points

 © 2014 Wiley Periodicals, Inc.

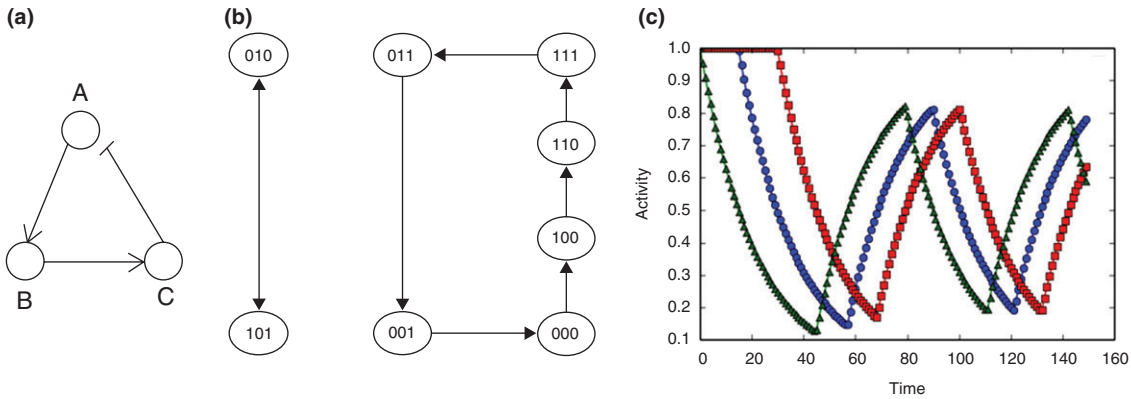


FIGURE 6 | Illustration of the dynamics of a piecewise linear model. (a) The directed network associated with the model forms a negative feedback loop. The Boolean update functions are completely determined by the network. (b) The state transition graph of a synchronous Boolean model of the network. The symbols correspond to the states of the system indicated in the order ABC. The Boolean model has two limit cycle attractors (sustained oscillations). (c) The time-course of the continuous variables associated with nodes A (green triangles), B (blue circles), and C (red squares) according to a piecewise linear model with $\gamma = 1$ and $\theta = 0.5$. The sustained oscillations agree with the six-state limit cycle of the synchronous Boolean model.

in the iterative numerical solution of the differential equations of the continuous model; typically yielding steady states that are close to the Boolean model's. This model has been applied to the regulatory network of helper T cells and has reproduced the molecular profiles corresponding to several helper T cell types. Another software, ODEfy,[111] converts Boolean regulation into Hill functions by multivariate polynomial interpolation. This transformation preserves the Boolean model's steady states. ODEfy models can be exported to other software such as the MATLAB Systems Biology Toolbox.[112]

Multistate discrete dynamic models can also be constructed to model biological systems. For example, in a three-state model, nodes can be assigned three states (e.g., −1, 0, 1 or 0, 1, 2) to represent under-activity (downregulation), normal activity, and over-activity (upregulation). As with Boolean models, truth tables can be constructed to represent the regulatory relationships among nodes. The future value of a regulated node will depend on the logical constraints designated by the modeler in the truth tables. Multiple alternative mathematical formalisms exist for a more compact representation of multistate discrete models: logical models,[113] implemented in the software GINSim[114] and polynomial dynamical systems[115] implemented in the software ADAM.[116] The logical functions used in the logical model framework specify the conditions for which the regulated node's state is different from the baseline.[113] Polynomial dynamical systems represent each truth table by a polynomial function. Polynomial algebra can then be used to identify the steady states of the model. One disadvantage of this method is that the polynomial representation (even for Boolean models) is less intuitive than logical models.[115]

CONCLUSION

Computational modeling of biological processes plays an important role in systems biology and enables efficient *in silico* experiments whose predictions greatly improve the design of wet-bench experiments. Although Boolean network models have a limited capacity to describe the quantitative characteristics of dynamic systems, they do exhibit considerable dynamic richness and were proven effective in describing the qualitative behaviors of biological systems. In addition, Boolean models were proven successful in predicting the key components of signal transduction and gene regulatory networks, and in proposing effective intervention strategies. The fact that Boolean models do not require the knowledge of kinetic parameters makes them a practical choice as well for systems where these parameters have not been measured. Thus Boolean models pass the two key tests: they are useful,[117] and they increase the understanding of the systems for which they were formulated.[2] The success of Boolean models illustrates that in at least a subset of biological systems the organization of network structure plays a more important role than the kinetic details of the individual interactions.[20,47] This inference can be fruitful in the context of understanding the functional and evolutionary constraints of biomolecular networks, and indeed Boolean models have been used to investigate these issues.[118,119]

 © 2014 Wiley Periodicals, Inc.

In practice, qualitative and quantitative models are complementary. The choice between qualitative models like Boolean network models and quantitative models described by differential equations depends on the availability of kinetic information, the size of the systems, and the types of questions to be addressed. Boolean networks can serve as a foundation of modeling regulatory and signaling networks on which more detailed continuous models can be built as kinetic information and quantitative experimental data become available.[120] Constructing first a Boolean and then a continuous or multistate model of the same system is often insightful. Several examples indicate that the different modeling approaches give rise to consistent results, e.g., the same qualitative set of attractors.[91,121,122] The simpler model can be used for efficient exploratory analysis to fix the model's structure and to help develop the refined continuous model, which in turn can be compared with quantitative biological observations.

REFERENCES

1. Papin JA, Hunter T, Palsson BO, Subramaniam S. Reconstruction of cellular signalling networks and analysis of their properties. *Nat Rev Mol Cell Biol* 2005, 6:99–111.
2. Lander AD. The edges of understanding. *BMC Biol* 2010, 8:40.
3. Albert R, Barabasi AL. Statistical mechanics of complex networks. *Rev Mod Phys* 2002, 74:47–97.
4. Kestler HA, Wawra C, Kracher B, Kuhl M. Network modeling of signal transduction: establishing the global view. *Bioessays* 2008, 30:1110–1125.
5. Karlebach G, Shamir R. Modelling and analysis of gene regulatory networks. *Nat Rev Mol Cell Biol* 2008, 9:770–780.
6. Pellegrini M, Haynor D, Johnson JM. Protein interaction networks. *Expert Rev Proteomics* 2004, 1:239–249.
7. Kauffman SA. Metabolic stability and epigenesis in randomly constructed genetic nets. *J Theor Biol* 1969, 22:437–467.
8. Thomas R. Boolean formalization of genetic control circuits. *J Theor Biol* 1973, 42:563–585.
9. Assmann SM, Albert R. Discrete dynamic modeling with asynchronous update, or how to model complex systems in the absence of quantitative information. *Methods Mol Biol* 2009, 553:207–225.
10. Chaves M, Sontag ED, Albert R. Methods of robustness analysis for Boolean models of gene control networks. *Syst Biol (Stevenage)* 2006, 153:154–167.
11. Giacomantonio CE, Goodhill GJ. A Boolean model of the gene regulatory network underlying Mammalian cortical area development. *PLoS Comput Biol* 2010, 6:e1000936.
12. Mendoza L, Alvarez-Buylla ER. Dynamics of the genetic regulatory network for *Arabidopsis thaliana* flower morphogenesis. *J Theor Biol* 1998, 193:307–319.
13. Bhardwaj G, Wells CP, Albert R, van Rossum DB, Patterson RL. Exploring phospholipase C-coupled Ca(2+) signalling networks using Boolean modelling. *IET Syst Biol* 2011, 5:174–184.
14. Saez-Rodriguez J, Simeoni L, Lindquist JA, Hemenway R, Bommhardt U, Arndt B, Haus UU, Weismantel R, Gilles ED, Klamt S, et al. A logical model provides insights into T cell receptor signaling. *PLoS Comput Biol* 2007, 3:e163.
15. Thakar J, Pilione M, Kirimanjeswara G, Harvill ET, Albert R. Modeling systems-level regulation of host immune responses. *PLoS Comput Biol* 2007, 3:e109.
16. Thakar J, Pathak AK, Murphy L, Albert R, Cattadori IM. Network model of immune responses reveals key effectors to single and co-infection dynamics by a respiratory bacterium and a gastrointestinal helminth. *PLoS Comput Biol* 2012, 8:e1002345.
17. Walsh ER, Thakar J, Stokes K, Huang F, Albert R, August A. Computational and experimental analysis reveals a requirement for eosinophil-derived IL-13 for the development of allergic airway responses in C57BL/6 mice. *J Immunol* 2011, 186:2936–2949.
18. Campbell C, Yang S, Albert R, Shea K. A network model for plant-pollinator community assembly. *Proc Natl Acad Sci USA* 2011, 108:197–202.
19. Davidich MI, Bornholdt S. Boolean network model predicts cell cycle sequence of fission yeast. *PLoS One* 2008, 3:e1672.
20. Li FT, Long T, Lu Y, Ouyang Q, Tang C. The yeast cell-cycle network is robustly designed. *Proc Natl Acad Sci USA* 2004, 101:4781–4786.
21. Christensen TS, Oliveira AP, Nielsen J. Reconstruction and logical modeling of glucose repression signaling pathways in *Saccharomyces cerevisiae*. *BMC Syst Biol* 2009, 3:7.
22. Mendoza L, Alvarez-Buylla ER. Genetic regulation of root hair development in *Arabidopsis thaliana*: a network model. *J Theor Biol* 2000, 204:311–326.
23. Li S, Assmann SM, Albert R. Predicting essential components of signal transduction networks: a dynamic model of guard cell abscisic acid signaling. *PLoS Biol* 2006, 4:e312.

 © 2014 Wiley Periodicals, Inc.

24. Akman OE, Watterson S, Parton A, Binns N, Millar AJ, Ghazal P. Digital clocks: simple Boolean models can quantitatively describe circadian systems. *J R Soc Interface* 2012, 9:2365–2382.
25. Chaves M, Albert R. Studying the effect of cell division on expression patterns of the segment polarity genes. *J R Soc Interface* 2008, 5(Suppl 1):S71–S84.
26. Zhang R, Shah MV, Yang J, Nyland SB, Liu X, Yun JK, Albert R, Loughran TP Jr. Network model of survival signaling in large granular lymphocyte leukemia. *Proc Natl Acad Sci USA* 2008, 105:16308–16313.
27. Saadatpour A, Wang RS, Liao A, Liu X, Loughran TP, Albert I, Albert R. Dynamical and structural analysis of a T cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia. *PLoS Comput Biol* 2011, 7:e1002267.
28. Schlatter R, Philippi N, Wangorsch G, Pick R, Sawodny O, Borner C, Timmer J, Ederer M, Dandekar T. Integration of Boolean models exemplified on hepatocyte signal transduction. *Brief Bioinform* 2011, 13:365–376.
29. Choi M, Shi J, Jung SH, Chen X, Cho KH. Attractor landscape analysis reveals feedback loops in the p53 network that control the cellular response to DNA damage. *Sci Signal* 2012, 5:ra83.
30. Sobie EA, Lee YS, Jenkins SL, Iyengar R. Systems biology—biomedical modeling. *Sci Signal* 2011, 4:tr2.
31. Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK. Physicochemical modelling of cell signalling pathways. *Nat Cell Biol* 2006, 8:1195–1203.
32. Kauffman SA. *The Origins of Order: Self Organization and Selection in Evolution*. New York: Oxford University Press; 1993.
33. Thomas R, d'Ari R. *Biological Feedback*. Boca Raton, FL: CRC Press; 1990.
34. Aldridge BB, Saez-Rodriguez J, Muhlich JL, Sorger PK, Lauffenburger DA. Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling. *PLoS Comput Biol* 2009, 5:e1000340.
35. Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA. Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. *PLoS Comput Biol* 2011, 7:e1001099.
36. Chaouiya C. Petri net modelling of biological networks. *Brief Bioinform* 2007, 8:210–219.
37. Ruths D, Muller M, Tseng JT, Nakhleh L, Ram PT. The signaling Petri net-based simulator: a non-parametric strategy for characterizing the dynamics of cell-specific signaling networks. *PLoS Comput Biol* 2008, 4:e1000005.
38. Samaga R, Saez-Rodriguez J, Alexopoulos LG, Sorger PK, Klamt S. The logic of EGFR/ErbB signaling: theoretical properties and analysis of high-throughput data. *PLoS Comput Biol* 2009, 5:e1000438.
39. Schlatter R, Schmich K, Avalos Vizcarra I, Scheurich P, Sauter T, Borner C, Ederer M, Merfort I, Sawodny O. ON/OFF and beyond--a boolean model of apoptosis. *PLoS Comput Biol* 2009, 5:e1000595.
40. Davidich M, Bornholdt S. The transition from differential equations to Boolean networks: a case study in modeling regulatory networks. *J Theor Biol* 2008, 225:269–277.
41. von Dassow G, Meir E, Munro EM, Odell GM. The segment polarity network is a robust developmental module. *Nature* 2000, 406:188–192.
42. Glass L, Kauffman SA. Logical analysis of continuous, nonlinear biochemical control networks. *J Theor Biol* 1973, 39:103–129.
43. Huang S, Eichler G, Bar-Yam Y, Ingber DE. Cell fates as high-dimensional attractor states of a complex gene regulatory network. *Phys Rev Lett* 2005, 94:128701.
44. Saadatpour A, Albert I, Albert R. Attractor analysis of asynchronous Boolean models of signal transduction networks. *J Theor Biol* 2010, 4:641–656.
45. Tyson JJ, Chen KC, Novak B. Sniffers, buzzers, toggles and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr Opin Cell Biol* 2003, 15:221–231.
46. Serra R, Villani M, Barbieri A, Kauffman SA, Colacci A. On the dynamics of random Boolean networks subject to noise: attractors, ergodic sets and cell types. *J Theor Biol* 2010, 265:185–193.
47. Albert R, Othmer HG. The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster*. *J Theor Biol* 2003, 223:1–18.
48. Mendoza L. A network model for the control of the differentiation process in Th cells. *Biosystems* 2006, 84:101–114.
49. Mendoza L, Pardo F. A robust model to describe the differentiation of T-helper cells. *Theory Biosci* 2010, 129:283–293.
50. Kaech SM, Hemby S, Kersh E, Ahmed R. Molecular and functional profiling of memory CD8 T cell differentiation. *Cell* 2002, 111:837–851.
51. Lund RJ, Loytomaki M, Naumanen T, Dixon C, Chen Z, Ahlfors H, Tuomela S, Tahvanainen J, Scheinin J, Henttinen T, et al. Genome-wide identification of novel genes involved in early Th1 and Th2 cell differentiation. *J Immunol* 2007, 178:3648–3660.
52. Mathur D, Danford TW, Boyer LA, Young RA, Gifford DK, Jaenisch R. Analysis of the mouse embryonic stem cell regulatory networks obtained by ChIP-chip and ChIP-PET. *Genome Biol* 2008, 9:R126.
53. Albert R, Assmann SM. Discrete dynamic modeling with asynchronous update or, how to model complex systems in the absence of quantitative information. In: Belostotsky D, Totowa NJ, eds. *Plant Systems Biology*, vol. 553. New York: Humana Press; 2009, 207–225.

 © 2014 Wiley Periodicals, Inc.

54. Albert R, DasGupta B, Dondi R, Kachalo S, Sontag E, Zelikovsky A, Westbrooks K. A novel method for signal transduction network inference from indirect experimental evidence. *J Comput Biol* 2007, 14:927–949.
55. Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, Kasif S, Collins JJ, Gardner TS. Large-scale mapping and validation of *Escherichia coli* transcriptional regulation from a compendium of expression profiles. *PLoS Biol* 2007, 5:e8.
56. Friedman N, Linial M, Nachman I, Pe'er D. Using Bayesian networks to analyze expression data. *J Comput Biol* 2000, 7:601–620.
57. Gardner TS, di Bernardo D, Lorenz D, Collins JJ. Inferring genetic networks and identifying compound mode of action via expression profiling. *Science* 2003, 301:102–105.
58. Akutsu T, Miyano S, Kuhara S. Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. *Pac Symp Biocomput* 1999, 4:17–28.
59. Ideker TE, Thorsson V, Karp RM. Discovery of regulatory interactions through perturbation: inference and experimental design. *Pac Symp Biocomput* 2000, 5:305–316.
60. Shmulevich I, Dougherty ER, Kim S, Zhang W. Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* 2002, 18:261–274.
61. Kim SY, Imoto S, Miyano S. Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Brief Bioinform* 2003, 4:228–235.
62. Kachalo S, Zhang R, Sontag E, Albert R, DasGupta B. NET-SYNTHESIS: a software for synthesis, inference and simplification of signal transduction networks. *Bioinformatics* 2008, 24:293–295.
63. yEd Graph Editor. Available at: http://www.yworks.com/en/products_yed_about.html.
64. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003, 13:2498–2504.
65. Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T. Cytoscape 2.8: new features for data integration and network visualization. *Bioinformatics* 2011, 27:431–432.
66. Klamt S, Saez-Rodriguez J, Gilles ED. Structural and functional analysis of cellular networks with CellNetAnalyzer. *BMC Syst Biol* 2007, 1:2.
67. Klamt S, Saez-Rodriguez J, Lindquist JA, Simeoni L, Gilles ED. A methodology for the structural and functional analysis of signaling and regulatory networks. *BMC Bioinformatics* 2006, 7:56.
68. Hagberg A, Schult D, Swart P. Exploring network structure, dynamics, and function using NetworkX. In: Proceedings of the 7th Python in Science Conference (SciPy 2008), Pasadena, CA, 2008.
69. Batagelj V, Mrvar A. Pajek – analysis and visualization of large networks. In: Jünger M, Mutzel P, eds. *Graph Drawing Software*. Berlin: Springer; 2003, 77–103.
70. Luscombe NM, Babu MM, Yu H, Snyder M, Teichmann SA, Gerstein M. Genomic analysis of regulatory network dynamics reveals large topological changes. *Nature* 2004, 431:308–312.
71. Han JD, Bertin N, Hao T, Goldberg DS, Berriz GF, Zhang LV, Dupuy D, Walhout AJ, Cusick ME, Roth FP, et al. Evidence for dynamically organized modularity in the yeast protein-protein interaction network. *Nature* 2004, 430:88–93.
72. Kim PM, Lu LJ, Xia Y, Gerstein MB. Relating three-dimensional structures to protein networks provides evolutionary insights. *Science* 2006, 314:1938–1941.
73. Ma HW, Buer J, Zeng AP. Hierarchical structure and modules in the *Escherichia coli* transcriptional regulatory network revealed by a new top-down approach. *BMC Bioinformatics* 2004, 5:199.
74. Sanz J, Navarro J, Arbues A, Martin C, Marijuan PC, Moreno Y. The transcriptional regulatory network of *Mycobacterium tuberculosis*. *PLoS One* 2011, 6:e22178.
75. Ma HW, Zeng AP. The connectivity structure, giant strong component and centrality of metabolic networks. *Bioinformatics* 2003, 19:1423–1430.
76. Ma'ayan A, Jenkins SL, Neves S, Hasseldine A, Grace E, Dubin-Thaler B, Eungdamrong NJ, Weng G, Ram PT, Rice JJ, et al. Formation of regulatory patterns during signal propagation in a mammalian cellular network. *Science* 2005, 309:1078–1083.
77. Milo R, Shen-Orr S, Itzkovitz S, Kashtan N, Chklovskii D, Alon U. Network motifs: simple building blocks of complex networks. *Science* 2002, 298:824–827.
78. Shen-Orr SS, Milo R, Mangan S, Alon U. Network motifs in the transcriptional regulation network of *Escherichia coli*. *Nat Genet* 2002, 31:64–68.
79. Alon U. *An Introduction to Systems Biology: Design Principles of Biological Circuits*, vol. 10. Boca Raton, FL: Chapman & Hall/CRC; 2006.
80. Sontag E, Veliz-Cuba A, Laubenbacher R, Jarrah AS. The effect of negative feedback loops on the dynamics of Boolean networks. *Biophys J* 2008, 95:518–526.
81. Bornholdt S. Boolean network models of cellular regulation: prospects and limitations. *J R Soc Interface* 2008, 5(Suppl 1):S85–S94.
82. Shmulevich I, Dougherty ER. *Probabilistic Boolean Networks: The Modeling and Control of Gene Regulatory Networks*. New York: SIAM; 2010.
83. Berestovsky N, Nakhleh L. An evaluation of methods for inferring Boolean networks from time-series data. *PLoS One* 2013, 8:e66031.

 © 2014 Wiley Periodicals, Inc.

84. Cheng X, Sun M, Socolar JE. Autonomous Boolean modelling of developmental gene regulatory networks. *J R Soc Interface* 2013, 10:20120574.
85. Murrugarra D, Veliz-Cuba A, Aguilar B, Arat S, Laubenbacher R. Modeling stochasticity and variability in gene regulatory networks. *EURASIP J Bioinform Syst Biol* 2012, 2012:5.
86. Chaves M, Albert R, Sontag ED. Robustness and fragility of Boolean models for genetic regulatory networks. *J Theor Biol* 2005, 235:431–449.
87. Harvey I, Bossomaier T. Time out of joint: attractors in asynchronous random Boolean networks. In: Proceedings of the Fourth European Conference on Artificial Life, Cambridge, UK, 1997.
88. Faure A, Naldi A, Chaouiya C, Thieffry D. Dynamical analysis of a generic Boolean model for the control of the mammalian cell cycle. *Bioinformatics* 2006, 22:e124–e131.
89. Tournier L, Chaves M. Uncovering operational interactions in genetic networks using asynchronous Boolean dynamics. *J Theor Biol* 2009, 260:196–209.
90. Garg A, Di Cara A, Xenarios I, Mendoza L, De Micheli G. Synchronous versus asynchronous modeling of gene regulatory networks. *Bioinformatics* 2008, 24:1917–1925.
91. Albert I, Thakar J, Li S, Zhang R, Albert R. Boolean network simulations for life scientists. *Source Code Biol Med* 2008, 3:16.
92. Mussel C, Hopfensitz M, Kestler HA. BoolNet—an R package for generation, reconstruction and analysis of Boolean networks. *Bioinformatics* 2010, 26:1378–1380.
93. Dubrova E, Teslenko M. A SAT-based algorithm for finding attractors in synchronous Boolean networks. *IEEE/ACM Trans Comput Biol Bioinform* 2011, 8:1393–1399.
94. Skodawessely T, Klemm K. Finding attractors in asynchronous Boolean dynamics. *Adv Complex Syst* 2011, 14:439–449.
95. Berenguier D, Chaouiya C, Monteiro PT, Naldi A, Remy E, Thieffry D, Tichit L. Dynamical modeling and analysis of large cellular regulatory networks. *Chaos* 2013, 23:025114.
96. Bilke S, Sjunnesson F. Stability of the Kauffman model. *Phys Rev E* 2002, 65:016129.
97. Naldi A, Remy E, Thieffry D, Chaouiya C. Dynamically consistent reduction of logical regulatory graphs. *Theor Comput Sci* 2011, 412:2207–2218.
98. Richardson KA. Simplifying Boolean networks. *Adv Complex Syst* 2005, 8:365–381.
99. Veliz-Cuba A. Reduction of Boolean network models. *J Theor Biol* 2011, 289:167–172.
100. Saadatpour A, Albert R, Reluga TC. A reduction method for Boolean network models proven to conserve attractors. *SIAM J Appl Dyn Syst* 2013, 12:1997–2011.
101. Huang S, Ernberg I, Kauffman S. Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective. *Semin Cell Dev Biol* 2009, 20:869–876.
102. Darabos C, Tomassini M, Giacobini M. Dynamics of unperturbed and noisy generalized Boolean networks. *J Theor Biol* 2009, 260:531–544.
103. Samaga R, Von Kamp A, Klamt S. Computing combinatorial intervention strategies and failure modes in signaling networks. *J Comput Biol* 2010, 17:39–53.
104. Wang RS, Albert R. Elementary signaling modes predict the essentiality of signal transduction network components. *BMC Syst Biol* 2011, 5:44.
105. Wolpert L, Beddington R, Brockes J, Jessell T, Lawrence P, Meyerowitz E. *Principles of Development*. London: Current Biology Press; 1998.
106. Zanudo JG, Albert R. An effective network reduction approach to find the dynamical repertoire of discrete dynamic networks. *Chaos* 2013, 23:025111.
107. De Jong H, Gouze JL, Hernandez C, Page M, Sari T, Geiselmann J. Qualitative simulation of genetic regulatory networks using piecewise-linear models. *Bull Math Biol* 2004, 66:301–340.
108. Thakar J, Saadatpour-Moghaddam A, Harvill ET, Albert R. Constraint-based network model of pathogen-immune system interactions. *J R Soc Interface* 2009, 6:599–612.
109. de Jong H, Geiselmann J, Hernandez C, Page M. Genetic network analyzer: qualitative simulation of genetic regulatory networks. *Bioinformatics* 2003, 19:336–344.
110. Di Cara A, Garg A, De Micheli G, Xenarios I, Mendoza L. Dynamic simulation of regulatory networks using SQUAD. *BMC Bioinformatics* 2007, 8:462.
111. Krumsiek J, Polsterl S, Wittmann DM, Theis FJ. Odefy—from discrete to continuous models. *BMC Bioinformatics* 2010, 11:233.
112. Schmidt H, Jirstrand M. Systems biology toolbox for MATLAB: a computational platform for research in systems biology. *Bioinformatics* 2006, 22:514–515.
113. Thomas R. Regulatory networks seen as asynchronous automata—a logical description. *J Theor Biol* 1991, 153:1–23.
114. Naldi A, Berenguier D, Faure A, Lopez F, Thieffry D, Chaouiya C. Logical modelling of regulatory networks with GINsim 2.3. *Biosystems* 2009, 97:134–139.
115. Veliz-Cuba A, Jarrah AS, Laubenbacher R. Polynomial algebra of discrete models in systems biology. *Bioinformatics* 2010, 26:1637–1643.
116. Hinkelmann F, Brandon M, Guang B, McNeill R, Blekherman G, Veliz-Cuba A, Laubenbacher R. ADAM: analysis of discrete models of biological

 © 2014 Wiley Periodicals, Inc.

systems using computer algebra. *BMC Bioinformatics* 2011, 12:295.

117. Box GEP. Robustness in the strategy of scientific model building. In: Launer RL, Wilkinson GN, eds. *Robustness in Statistics*. New York: Academic Press; 1979, 201–236.

118. Payne JL, Wagner A. Constraint and contingency in multifunctional gene regulatory circuits. *PLoS Comput Biol* 2013, 9:e1003071.

119. Oikonomou P, Cluzel P. Effects of topology on network evolution. *Nat Phys* 2006, 2:532–536.

120. Wittmann DM, Krumsiek J, Saez-Rodriguez J, Lauffenburger DA, Klamt S, Theis FJ. Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling. *BMC Syst Biol* 2009, 3:98.

121. Chaves M, Preto M. Hierarchy of models: From qualitative to quantitative analysis of circadian rhythms in cyanobacteria. *Chaos* 2013, 23:025113.

122. Sun MY, Cheng XR, Socolar JES. Causal structure of oscillations in gene regulatory networks: Boolean analysis of ordinary differential equation attractors. *Chaos* 2013, 23.

FURTHER READING

Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger DA. Logic-based models for the analysis of cell signaling networks. *Biochemistry* 2010, 49:3216–3224.

Wynn ML, Consul N, Merajver SD, Schnell S. Logic-based models in systems biology: a predictive and parameter-free network analysis method. *Integr Biol* 2012, 4:1323–1337.

Thakar J, Albert R. Boolean models of within-host immune interactions. *Curr Opin Microbiol* 2010, 13:377–381.

 © 2014 Wiley Periodicals, Inc.