Boolean modeling in systems biology: an overview of methodology and applications

2012 Phys. Biol. 9 055001

(http://iopscience.iop.org/1478-3975/9/5/055001)

Phys. Biol. **9** (2012) 055001 (14pp) doi:10.1088/1478-3975/9/5/055001

Boolean modeling in systems biology: an overview of methodology and applications

Rui-Sheng Wang[1], **Assieh Saadatpour**[2] **and Réka Albert**[1,3]

[1] Department of Physics, The Pennsylvania State University, University Park, PA 16802, USA
[2] Department of Mathematics, The Pennsylvania State University, University Park, PA 16802, USA

E-mail: ralbert@phys.psu.edu

Received 6 January 2012
Accepted for publication 6 June 2012
Published 25 September 2012
Online at stacks.iop.org/PhysBio/9/055001

Abstract
Mathematical modeling of biological processes provides deep insights into complex cellular systems. While quantitative and continuous models such as differential equations have been widely used, their use is obstructed in systems wherein the knowledge of mechanistic details and kinetic parameters is scarce. On the other hand, a wealth of molecular level qualitative data on individual components and interactions can be obtained from the experimental literature and high-throughput technologies, making qualitative approaches such as Boolean network modeling extremely useful. In this paper, we build on our research to provide a methodology overview of Boolean modeling in systems biology, including Boolean dynamic modeling of cellular networks, attractor analysis of Boolean dynamic models, as well as inferring biological regulatory mechanisms from high-throughput data using Boolean models. We finally demonstrate how Boolean models can be applied to perform the structural analysis of cellular networks. This overview aims to acquaint life science researchers with the basic steps of Boolean modeling and its applications in several areas of systems biology.

1. Introduction

Systems biology aims to elucidate how complex behaviors of biological systems emerge from the properties of the components and interactions in the systems. It uses a combination of experimental techniques and computational approaches to gain global insights into complex biological systems. The experimental techniques employed in systems biology tend to have high-throughput capabilities, and are thus able to determine the abundance or activity of numerous components at the same time. For example, the abundance of mRNA transcripts of thousands of genes can be measured by microarrays [19] or RNA-seq [66]. Quantitative protein concentrations and post-translational modifications can be determined by proteomics and phosphoproteomics studies conducted through mass spectrometry (MS) [7069] or two-dimensional (2D) gels [32, 105]. Metabolomic profiles generated by gas chromatography (GC)-MS or liquid chromatography (LC)-MS can measure the composition and concentration of both targeted and untargeted metabolites [50, 54]. There are also high-throughput experiments that can detect the interactions among components such as protein–protein interactions [102], transcriptional regulations (protein–DNA interactions) [56, 39] and genetic interactions [95]. In addition to high-throughput assays, small-scale experiments that study fewer components and interactions involved in specific biological processes provide high-quality and reliable focused knowledge for biological systems [15, 51]. Taking drought-induced signaling in plants as an example, numerous small-scale experiments performed over decades generated an abundance of individual components and causal interactions that mediate this signaling in guard cells. Such information was then assembled into a guard cell abscisic acid (ABA) signal transduction network [51].

Experimental data from high-throughput technologies and small-scale studies provide a rich source for understanding the system-level mechanisms of biological processes [72]. However, a complement of computational and modeling approaches is necessary to obtain mechanistic insights from the data and generate testable hypotheses. Computational

[3] Author to whom any correspondence should be addressed.

methods that have been used in systems biology can be classified into top-down methods and bottom-up methods [88]. Top-down methods such as statistical analyses and static network models are applied to high-throughput omics data and aim to decipher the organization of the underlying systems and mine information specific to a biological process [88, 47, 5]. Methods from this class do not require kinetic parameters and can be applied to the analysis of genome-scale data with tens of thousands of components or interactions to obtain coarse-grained knowledge about biological systems. Bottom-up methods model how interacting components such as genes, proteins and metabolites achieve the dynamic behaviors of cellular systems. This class of methods usually starts with hypotheses of biological mechanisms generated from individual small-scale experiments. Continuous dynamic modeling, the most widely used bottom-up method [88, 10, 48], requires sufficient mechanistic details and kinetic parameters such as synthesis and degradation rates, making it practical on systems with only tens of components or less. Discrete dynamic modeling such as Boolean network models [9, 12, 91], multi-valued logical models [11, 64] and Petri nets [22, 76], does not require kinetic parameters and is able to provide qualitative dynamic descriptions of system behaviors. These approaches can be employed for systems with hundreds of components and have been increasingly used in modeling biological networks [81, 82, 84, 85].

Boolean network models, which were initially proposed as prototypical models of genetic regulatory networks [49, 94], are a special case of discrete dynamic models. A Boolean network consists of a set of nodes whose state is binary and is determined by other nodes in the network through Boolean functions. In terms of complexity, Boolean networks lie between static network models and continuous dynamic models [47], making them a tractable and powerful approach to modeling large-scale biological systems. After assembling individual components and regulatory interactions involved in a system into a coherent network representation, Boolean models can be used to describe the qualitative temporal behavior of the system and to understand how perturbations may alter its behaviors. They also lead to predictive testable hypotheses which are especially valuable in poorly understood large-scale systems [12, 38]. Boolean networks have been successfully applied in modeling many gene regulatory and signaling networks in a variety of organisms [59, 92, 103, 98, 8, 78], and have been reviewed in several survey articles [100, 65, 9, 12]. In addition, Boolean networks have been used as models in reverse engineering of biological networks, e.g., to infer regulatory interactions and signaling pathways from gene expression or proteomics data [57, 79, 80].

Here, we provide a methodology overview of Boolean modeling in systems biology and illustrate it using examples from our research. In section 2, we describe the main steps of dynamic modeling of cellular networks and their implementations, from construction of the network diagram to the generation of novel hypotheses. Section 3 presents the methodologies developed for a crucial modeling step, namely analysis of the possible long-term behaviors (attractors) of the system. Various applications of Boolean models show that they are useful both as a top-down method and as a bottom-up method in understanding system-level mechanisms of biological processes. Thus in section 4 we review methods for reverse engineering of biological regulatory networks using Boolean models. We end by demonstrating in section 5 how Boolean models can be employed for functional (input–output) analysis of cellular networks without the necessity of performing dynamic simulation.

2. Boolean dynamic modeling of cellular networks

A Boolean variable assumes only two values (0 and 1) corresponding to the logic values FALSE and TRUE. When representing the state of a biological entity by a binary variable, the two states are usually referred to as OFF and ON. A Boolean function with k variables is a mapping B: $\{0,1\}^k \rightarrow \{0,1\}$ from the set of all k-tuples over $\{0, 1\}$ to a binary output. This function describes how to determine a Boolean-valued output based on certain logical operations from k binary inputs [17]. The basic logical operations include AND, OR and NOT. For example, $D = (A$ OR $B)$ AND NOT C is a Boolean function with three variables. A Boolean function can also be represented by a truth table, wherein each row lists a combination of values of Boolean variables and its associated output value. The truth table of a Boolean function with k variables has 2^k rows and $k+1$ columns.

A Boolean network model consists of a set of Boolean variables $\{\sigma_1, \sigma_2, \ldots, \sigma_n\}$ whose value is determined by other variables in the network through a set of Boolean functions $B = \{B_1, B_2, \ldots, B_n\}$, one assigned to each variable. In a Boolean dynamic model, the value of each variable σ_i is determined by the current or prior values of its regulators (inputs), depending on the updating schemes used in the model. The synchronous scheme updates all variables in the model simultaneously, i.e., the value of each variable at time $t+1$ is determined by its k_i inputs at time t:

$$\sigma_i^{t+1} = B_i\left(\sigma_{i_1}^t, \sigma_{i_2}^t, \ldots, \sigma_{i_{k_i}}^t\right).$$

This update mode is deterministic. In asynchronous schemes, the variables are updated in a non-synchronous manner:

$$\sigma_i^* = B_i\left(\sigma_{i_1}, \sigma_{i_2}, \ldots, \sigma_{i_{k_i}}\right),$$

where the asterisk denotes the new value of the variable σ_i, $i = 1, 2, \ldots, n$. The values of its inputs on the right-hand side of the equation can be current or prior, depending on the individual timescales. As we describe in the following section, both deterministic and stochastic asynchronous schemes have been employed for modeling biological systems.

A Boolean network model can be projected to a directed graph $G(V, E)$, where the node set $V = \{v_1, v_2, \ldots, v_n\}$ corresponds to the Boolean variables, and the edge set E is implicitly defined by the Boolean functions in the model. Each edge has a sign implying whether the input node has a positive or negative effect on the target node. We note that Boolean models contain information additional to the directed and signed network diagram, since the same diagram can

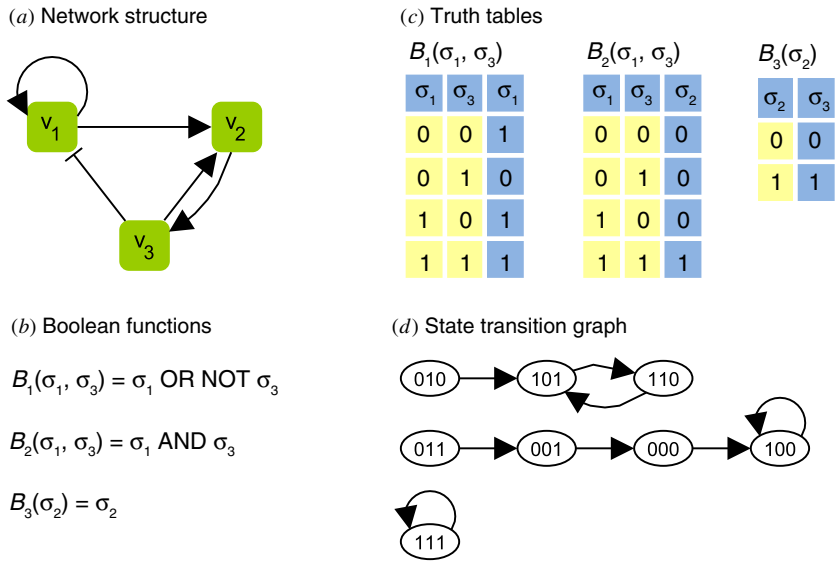

Figure 1. A simple Boolean network model. (*a*) The directed graph (network structure) associated with the Boolean model. The edges with sharp arrows represent positive effects and the edges with blunt arrows denote negative effects. Note that the graph does not uniquely determine the Boolean functions for nodes v_1 and v_2. (*b*) The Boolean functions in the model. Note that the graph in (*a*) can support alternative Boolean functions, specifically $B_1(\sigma_1, \sigma_3) = \sigma_1$ AND NOT σ_3, $B_2(\sigma_1, \sigma_3) = \sigma_1$ OR σ_3. (*c*) The truth tables of the Boolean functions given in (*b*). (*d*) The state transition graph of the Boolean model constructed using the synchronous updating scheme. The states 100 and 111 are the fixed points of the system, and the states 101 and 110 form a limit cycle.

correspond to several alternative Boolean functions and the Boolean model adds a dynamic layer to the network in the form of a state variable $\sigma_i(t)$ for each node v_i. The state of the system at time t can be represented by a vector $(\sigma_1(t), \sigma_2(t), \ldots, \sigma_n(t))$ with the ith element representing the state of node v_i at time t. All possible states of the system, a total of 2^n, make up its state space. The possible trajectories in the state space can be represented by a state transition graph, wherein nodes are states of the system and edges represent the allowed transitions among the states based on an updating scheme. By updating the nodes' states at each time step, the model evolves over time following a trajectory of states and eventually reaches a steady state (fixed point) or a set of recurring states. These steady or recurring states are collectively referred to as attractors. The set of initial states that leads the model to a specific attractor is called the basin of attraction of that attractor. Figure 1 illustrates a simple Boolean network and its associated Boolean functions, truth tables and state transition graph based on the synchronous scheme.

Boolean networks, as a special case of discrete dynamic modeling, provide an efficient formalism to describe the dynamics of biological systems [49, 94]. The directed and signed graph projection of a Boolean network model can be directly related to the pathway diagram of a biological regulatory system. Each node v_i in the Boolean network stands for a biological component such as a gene, protein, metabolite, an ion channel, or a stimulus (signal), which is associated with a binary state (expression level, concentration, or activity) σ_i. The state $\sigma_i = 1$ (ON) represents that component v_i is active or expressed, or has an above-threshold concentration, and $\sigma_i = 0$ (OFF) denotes that it is inactive or not expressed, or has a below-threshold concentration. The thresholds invoked in the definition of states do not need to be quantified, as long as it is known that a concentration level exists above which the component in question can effectively regulate its downstream targets. Each Boolean function represents the conditional dependence of input components in regulating the downstream target component. The parameter-free nature and qualitative features of Boolean modeling make it suitable for analyzing the complex behaviors of a large-scale system, such as the activity of components in a steady state, the activity changes of components following a perturbation, the input–output relations of the system and the stability of cellular responses to a signal.

Boolean dynamic modeling of biological networks entails six main steps, as shown in figure 2. In this overview, we provide a brief description of each of these steps in the following six paragraphs; a more detailed description can be found in [9, 12]. The first step is to *synthesize the network structure* by extensively collecting the relevant literature and experimental data concerning the biological system of interest. Although for many biological processes different experiments generate an abundance of relevant components and causal interactions, there is insufficient information on the overall structure and mechanisms of these processes. Therefore, information sources from individual experiments need to be assembled and integrated. Experimental evidence about the involvement of a component or a regulatory relationship in a biological process has several types. For example, the concentration change of a protein after treating the system with an input signal (ligand) indicates that this protein may be a component of this ligand's signal transduction network. Such evidence can be collected from high-throughput gene expression, proteomics and metabolomics data. In addition, if knocking out or over-expressing a component leads to changes in the relevant cellular response, it can be concluded that

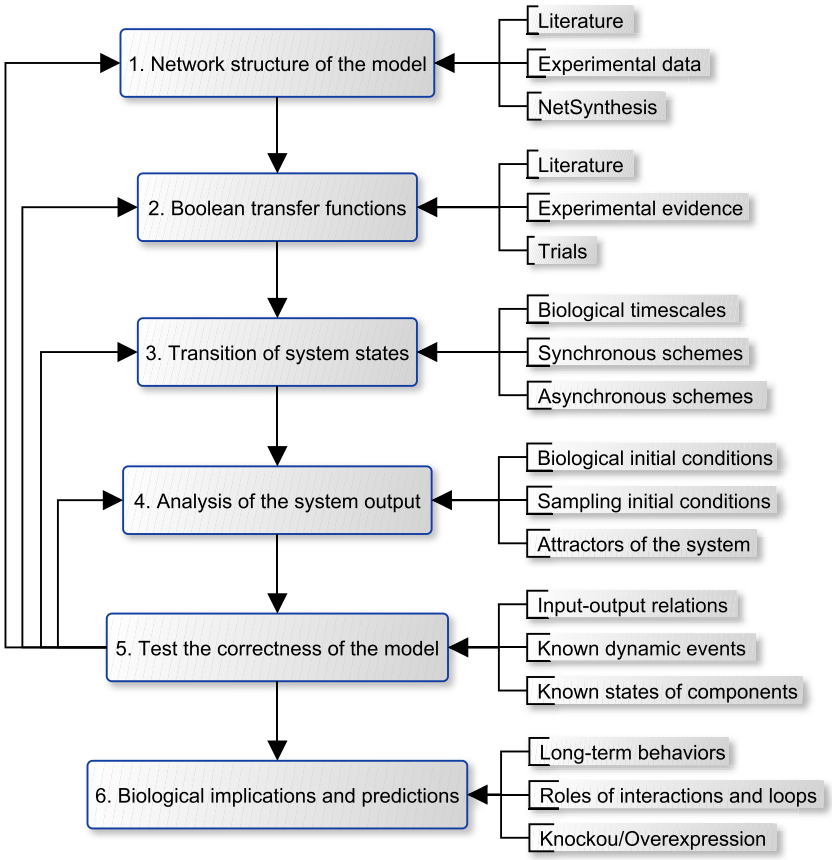

Figure 2. The main steps in Boolean dynamic modeling of biological systems.

this component is involved in the biological process. Causal relationships between components can be collected from high-throughput phosphoproteomics, protein–DNA interaction, as well as genetic interaction studies. These data sources may not be specific to the biological system of interest. To make them context-specific, one can complement them with direct biochemical evidence from small-scale experiments. If relevant information is sufficient, these causal relationships can be represented as directed edges from one component to another characterized by one of two signs: activating (positive) or inhibitory (negative). In some cases, genetic evidence from multiple experiments leads to composite causal relationships which can be broken down into component-to-component relationships depending on the concrete situation [6, 7]. Dealing with such cases is made simpler by the software NET-SYNTHESIS that finds the most parsimonious network incorporating all known components and the causal relationships between them [46, 6].

The network structure assembled in the first step significantly constrains, but does not uniquely determine, the dependency relationships among node states. Thus the second step is to *determine the Boolean transfer functions* based on evidence from the literature and experimental observations. Many biological events can be qualitatively represented by Boolean functions. For example, consider a protein P whose phosphorylated form regulates downstream processes (simply said, it is active). We can designate the state of protein P as 1 (ON) if it is predominantly in the phosphorylated form, and as 0 (OFF) if it is predominantly in the unphosphorylated form. If P is solely regulated by a kinase K that phosphorylates it, the Boolean transfer function for the state of P can be written as $P^* = K$, where for simplicity the node states are denoted by the node names, and the asterisk refers to the new state of P. If instead of a kinase the protein P interacts with and is dephosphorylated by a phosphatase R, the transfer function for P can be written as $P^* = \text{NOT R}$. In many cases, the activation of a component requires two or more regulators. For example, the transcription process of a gene G may require 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: $G^* = \text{P1 AND P2}$. If a component is regulated by multiple regulators and any of them can activate the component independently, e.g., a protein with multiple phosphorylation sites, the independent effects of these regulators on the target component can be captured by the OR operator. The Boolean transfer function for a component can be a complicated combination of AND, OR and NOT operations. If the transfer function for a component is not fully known, several variants can be tried by comparing their dynamic sequences with observations for the real system [18].

The dynamics of a Boolean model stems from the *transition of one system state to another*, and is determined by the transfer functions and influenced by the chosen updating scheme. The synchronous scheme is the simplest update mode, wherein the states of all nodes are updated simultaneously according to the last state of the system.

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. The state trajectory of the system is deterministic under the synchronous scheme and any system state can have at most one successor in the state transition graph. However, cellular systems are complicated, and most often the timescales of biological events are different and can vary widely from fractions of seconds to hours [72]. The synchronous scheme cannot account for such variations. In asynchronous models, the states of the nodes are updated in a non-synchronous manner, depending on the timescales of individual biological events. There are deterministic asynchronous schemes with fixed individual timescales [24] and stochastic asynchronous schemes wherein all nodes are updated according to a random order, or one node is randomly selected to be updated [23, 40]. In stochastic asynchronous models, the same initial condition can lead to different successors in the state transition graph due to the randomness involved in the update scheme. Asynchronous schemes can be informed by knowledge about the timescales of some components if such knowledge is available. Updating schemes have a considerable effect on the dynamics of a system. One can choose a scheme that is most realistic for the biological system of interest, or compare different schemes on the same system [77].

Starting from an initial condition the model evolves over time by transitioning from one state to another, and finally stabilizes in an attractor representing the long-term behavior of the system. The attractors of regulatory and signaling networks usually correspond to the steady activation states of components or to cellular phenotypes [58, 78, 33]. Therefore, *identifying the possible attractors* is useful and biologically relevant. Particularly, it allows one to examine the activities of components in a steady cellular state and compare them with experimental observations. It also helps to determine critical components for cellular phenotypes by examining the changes in the system's attractors if a certain component is knocked out (fixed in the OFF state) or over-expressed (fixed in the ON state). Different initial conditions may lead the system to different attractors. Ideally, one can start from a biologically relevant initial condition if it is known *a priori*. If the available information is insufficient, one can sample a large number of initial conditions and calculate the fraction of realizations of a certain attractor, representing the probability that the system attains the corresponding cellular phenotype [59, 103]. As a Boolean network model with n nodes has 2^n possible initial conditions, detecting attractors of a large system under synchronous and asynchronous schemes is a challenging problem; for this reason we dedicate section 3 to its discussion.

An important step of Boolean dynamic modeling of biological systems is to *test the correctness of the model*. There are several ways to validate certain features of the model. For example, the model must be able to reproduce prior experimental observations such as input–output relations, dynamic behaviors and cellular responses. If the model fails to do this, one needs to go back and check whether some important components or interactions are missing from the network structure, or whether some Boolean transfer functions are incomplete or wrong (e.g., use AND instead of OR or vice versa). The failure may also arise from the improperness of updating schemes or initial conditions. After several rounds of iterations, a Boolean dynamic model consistent with all known experimental observations can be obtained. An advanced strategy for validating the model is to conduct new experiments on the activity of some components. However such experiments can be very time-consuming and may not be the best investment in the model-testing stage. It has been revealed that many biological systems are robust [97, 8, 58], 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, rewiring a pair of interactions, adding or deleting a component or interaction. A good model can accommodate the majority of small perturbations, which reflects the adaptability of the system under diverse circumstances.

The first five steps are time- and labor-intensive, but all the effort is well worth it when the fragmented knowledge from individual experiments and data sources is assembled and integrated into a predictive model. The power of Boolean dynamic modeling is its ability to predict the outcomes 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 [78, 33]. 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. The outcomes of system perturbations (e.g., knockout or over-expression of certain components) can predict the changes in the steady-state activity of components and identify essential components accounting for phenotype traits [103, 92, 78, 59]. We can also predict the biological role of regulatory interactions and feedback loops by removing them and comparing the dynamic sequences before and after the perturbations. In summary, the model not only provides a system-level picture to understand the underlying mechanisms of the biological process, but also can direct follow-up targeted experiments and save the cost of exploratory wet-bench analysis.

Boolean networks have been successfully applied in modeling gene regulatory and signaling networks in a variety of biological systems [59, 34, 85, 8, 81, 82, 65, 100]. They have also been used to analyze human signaling networks associated with diseases to predict pathogenesis mechanisms and potential therapeutic targets [103, 78, 84, 92, 91, 98]. Many software tools are available for Boolean dynamic modeling of biological systems, such as BooleanNet [4], BoolNet [67], SimBoolNet [106] and ChemChains [42]. Several software packages also support multi-valued logical dynamic modeling, such as GINsim [36], SQUAD [28] and ADAM [43]. In addition to logic operation-based Boolean networks, threshold Boolean networks have been used in modeling biological networks at both cellular and population levels [58, 20]. Piecewise linear models are a hybrid of Boolean models and differential equation-based continuous models [35], and have been fruitfully applied due to their attractive combination

of continuous time, quantitative information and few kinetic parameters [24, 27, 93]. The methodologies and modeling steps described above apply to threshold Boolean models and piecewise linear models as well. In particular, one can utilize the software packages Genetic Network Analyzer [26] or BooleanNet [4] for qualitative modeling of biological networks based on piecewise linear models.

3. Attractor analysis of Boolean dynamic models

As it has been mentioned before, the attractors of a system represent the long-term behavior of the system. The simplest type of attractor is a single state called a fixed point (steady state), which remains unchanged under additional updates of the system. Since fixed points are time-independent, the fixed-point repertoire of a Boolean model is the same regardless of the manner of update (synchronous or asynchronous). We note, however, that the choice of updating scheme can affect the probability with which the system reaches these fixed points when starting from a given initial condition. In addition to fixed points, complex attractors in which the system oscillates among a set of states may appear in the state transition graphs of the system. The complex attractors of deterministic and stochastic Boolean models can be different. In synchronous and deterministic asynchronous models, the system oscillates regularly among the states in a complex attractor, which in this case is referred to as a limit cycle. The number of states in the limit cycle is called the period or length of the attractor. In stochastic asynchronous models, on the other hand, the system oscillates irregularly among a set of states; these complex attractors are alternatively referred to as loose attractors [40]. It was observed that limit cycles present in synchronous Boolean models can be absent from the corresponding asynchronous Boolean models [31, 77].

For small Boolean network models, the fixed points can be found by analytical methods. For example, taking away the time dependence of the transfer functions in a Boolean model, they form a system of time-independent Boolean equations described by $B_i(\sigma_1, \ldots, \sigma_n) = \sigma_i$, $i = 1, 2, \ldots, n$, where n is the number of nodes in the network. All the possible solutions of this system correspond to the fixed points of the Boolean model. For example, the fixed points of the simple network in figure 1 can be obtained analytically by solving the following system of equations: $\sigma_1 = \sigma_1$ OR NOT σ_3, $\sigma_2 = \sigma_1$ AND σ_3, and $\sigma_3 = \sigma_2$. Substituting the last equation into the second one results in $\sigma_3 = \sigma_1$ AND σ_3. Substituting this equation into the first one and simplifying the resulting equation using Boolean algebra leads to $\sigma_1 = 1$. As a result, $\sigma_3 = (1 \text{ AND } \sigma_3) = \sigma_3$, that is, σ_3 can be either 0 or 1. Noting that $\sigma_2 = \sigma_3$, we obtain two fixed points of the system, 100 and 111. Boolean models of signaling networks with one or more steady signals (inputs) usually contain nodes whose states stabilize after a transient period irrespective of the update methods or initial conditions of other nodes. The logical steady state analysis proposed in [53] can find (partial) fixed points of Boolean models of signaling networks by propagating the input signals to the output layer. In addition, methods based on scalar equations and reduced scalar equations [30, 41] can be used to obtain limit cycles of synchronous models analytically. These equations, which are ordinary recurrence equations for the nodes of a Boolean network, are obtained by iterating the original Boolean rules. Unfortunately, such techniques are not practical for large Boolean networks [104]. Finding loose attractors of stochastic asynchronous models analytically is difficult even for small networks since the state transitions involve stochasticity.

In addition to analytical methods, another class of approaches to finding attractors takes advantage of the properties of state transition graphs. We note that fixed points in a state transition graph are nodes without outgoing edges except self-loops. A limit cycle in the state transition graph of a synchronous or deterministic asynchronous model is a set of nodes that make up a cycle without outgoing edges. For example, in figure 1(*d*), the states 100 and 111 are the two fixed points of the system, and the states 101 and 110 form a limit cycle. A loose attractor in the state transition graph of a stochastic asynchronous model is a set of states that form a strongly connected component without outgoing edges. The attractors of small Boolean networks can be found by constructing complete state transition graphs using numerical simulations and then searching such graphs. However, constructing the state transition graph of a relatively large Boolean network is usually time-consuming. Search methods that utilize the special features of attractors in state transition graphs without the necessity of checking all possible trajectories have also been developed. Dubrova and Teslenko [29] proposed a SAT-based bounded model checking algorithm for finding all attractors in synchronous Boolean networks. This method first uses propositional formulae to represent the state transitions in l time steps, i.e., the paths of length l in the state transition graph. Then it checks if such paths exist and contain cycles so as to determine whether all attractors have been already identified or the search should continue by increasing the path length. To combat the state space explosion problem, Garg *et al* [33] used reduced ordered binary decision diagrams (BDDs) to represent Boolean functions and developed an algorithm to identify all attractors in synchronous Boolean networks by computing forward and backward reachable sets of seed states. They further extended this algorithm to identify loose attractors in a special case of asynchronous Boolean models by introducing a combined synchronous–asynchronous traversal technique to search reduced ordered BDDs [33]. Recently, Skodawessely and Klemm [87] proposed a method to identify attractors of asynchronous Boolean models based on a systematic removal of state transitions by stabilizing certain nodes' state. This method reduces the state transition graph into an acyclic graph such that all attractors become fixed points of this graph and can be enumerated with little effort in most instances. The attractors of the original dynamics model, each containing at least one fixed point of the reduced model, can be found by depth-first search seeded at each fixed point of the reduced model [87].

Identifying attractors of large-scale Boolean networks is a computationally difficult problem for both synchronous and asynchronous models [104]. One way to overcome

this difficulty is to simplify the networks prior to dynamic analysis. To this end, several reduction techniques have been proposed [16, 68, 77, 96, 75]. Certain variables in a Boolean network model evolve to the same steady state independent of the initial condition and thus are not relevant to the task of attractor identification. Such stable variables can be found by inspection of transition functions and network connectivity. A network reduction technique based on the removal of frozen nodes (stable variables) and network leaves (i.e. nodes with no outgoing edges) has been used to simplify random Boolean networks [16, 75]. In [16], the frozen nodes were identified in a way similar to the logical steady state analysis, whereas in [75], first a random sampling method for the initial states was used to determine a subset of the attractors, and then a minimum set of frozen nodes was found by identifying the nodes whose state was the same in all attractors. The reduction method proposed in [68, 96] consists of iteratively removing nodes without a self-loop from the network and simplifying the redundant transfer functions. This method was proven to preserve the fixed points of a system, but it may introduce spurious oscillations into the reduced model [68].

In [77], we proposed a two-step reduction method for signal transduction networks, which first identifies and removes stabilized nodes, and then iteratively removes simple mediator nodes (e.g., nodes with one incoming edge and one outgoing edge). To identify the stabilized nodes, one needs to first fix the known state of signal nodes, and iteratively identify the nodes whose rules depend on the signal and/or already-stabilized nodes and simplify those rules based on Boolean algebra. This procedure results in identification of either the fixed point(s) of the system, or the partial fixed point(s) and a remaining system of equations. We note that the first step of this reduction method is similar to the logical steady state analysis in [53]. If this system of equations is small enough, one can obtain the attractors of the reduced network by using the methods mentioned before. Otherwise, one can merge simple mediator nodes and obtain a simpler set of equations whose attractors can be easily identified by the methods suitable for small networks. The attractors of the reduced network can be readily extended to those of the whole system by considering the states of the stabilized nodes and the removed simple mediator nodes. This reduction method facilitates the identification of attractors for large networks. We employed it to identify attractors of the ABA signal transduction network [77] as well as a signaling network corresponding to the disease T-LGL leukemia [78]. For example, for the former, the properties of the state transition graph, including attractors of the system and their basins of attraction, were compared in a synchronous as well as three different asynchronous models [77].

Once the attractors of the Boolean model of a system are determined, the activity of components in relevant cellular responses or phenotype traits can be predicted. For example, the Boolean dynamic model of a T-helper (Th) cell differentiation network [33] has three fixed points, representing the activation patterns of components observed in Th0, Th1 and Th2 cells. For the T-LGL leukemia survival signaling network, the attractors of the system representing the healthy and disease conditions were identified, and the latter successfully unraveled the T-LGL (disease) states of all the components in the network, including the ones that were experimentally undocumented before [78]. Having the attractors of a system, one can also perform dynamic perturbation analysis by knocking out the nodes that stabilize at ON and over-expressing the ones that stabilize at OFF in an attractor. This allows identifying component manipulations that can change the ultimate outcome of the system. 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 in agreement with the available experimental data and the rest can guide future experiments [78].

It is worth noting that a stochastic asynchronous Boolean model can be described by a discrete-time Markov chain process [23, 34, 77, 78]. The transition matrix P of the Markov chain contains the probabilities of transition from each state to all other states in the state space which can be obtained according to the asynchronous scheme used in the model. The ijth entry $p_{ij}^{(m)}$ of the matrix P^m gives the probability that the system will be in state s_j after m time steps when starting from state s_i. Let $\pi(0)$ be the probability vector which represents the initial distribution of states of the system. Then $\pi(m) = \pi(0)P^m$ gives the probability that the system is in each state after m time steps. In particular, this formula gives the probability that the system ends up in different steady states or in transient states in loose attractors if m is large enough [34]. When the asynchronous model has at least one fixed point, and it is possible to reach a fixed point from each transient state, it corresponds to an absorbing Markov chain [37]. In such Markov chains, starting from any transient state s_i of the system one can obtain the absorption time t_i (the expected number of time steps to reach any of the fixed points) as well as the absorption probabilities u_{ij} (the probabilities of reaching each fixed point s_j from the transient state s_i) [37]. The former can be calculated using the formula $T = (I-Q)^{-1}O$, where I is an identity matrix, O is a column vector all of whose elements are 1 and Q is the sub-matrix of P containing the transition probabilities from each transient state to all other transient states. The latter can be obtained by the formula $U = (I-Q)^{-1}R$, where R is the sub-matrix of P containing the transition probabilities from each transient state to the fixed points. Calculating absorption probabilities is particularly important for the systems with multiple fixed points as it can provide the probabilities of the system being in each steady state. For example, in the T-LGL leukemia signaling network, the analysis of the absorption probabilities revealed that for the majority of the states that can reach both the healthy and disease fixed points, the probabilities of reaching the healthy fixed point is higher than those of the disease fixed point [78].

4. Inferring biological mechanisms with Boolean models

As we described in section 2, in Boolean dynamic modeling of biological systems, individual regulatory relationships,

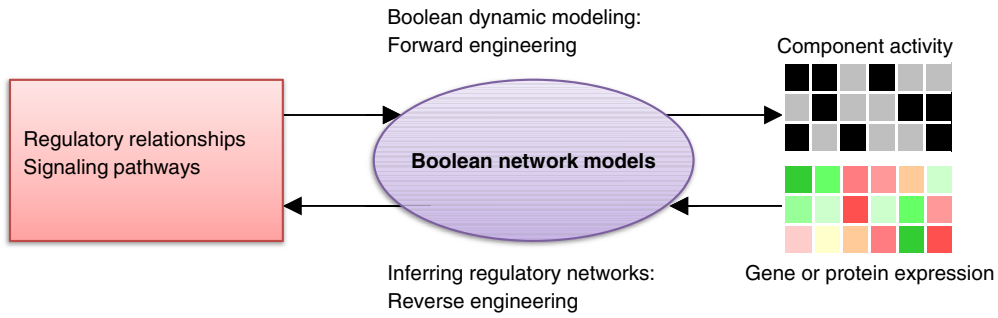

Figure 3. Applications of Boolean network models in forward and reverse engineering of biological networks.

partial signaling pathways and conditional dependence among regulators are assembled from experimental observations and integrated together to form a dynamic model of a relatively complete regulatory network. This dynamic model can predict the qualitative activity of components such as proteins and metabolites. The bottom-up approach used to construct the dynamic model is sometimes referred to as forward engineering of biological regulatory networks [25, 73]. On the other hand, high-throughput techniques can generate the abundance profiles of genes, proteins and metabolites in a steady or transient state of the cell. These measurements reflect the quantitative activities of these biological components in the corresponding biological processes which arise from the system-level dynamic regulation of various molecules. Reverse engineering of biological networks, as shown in figure 3, aims to recover these underlying regulatory relationships or signaling pathways from observed activity data [13, 14, 89, 79].

Boolean networks are one of the most popular models used in reverse engineering of biological regulatory networks, wherein the objective is to infer not only the network structure but also the Boolean transfer functions from experimental abundance profiles. An early study in this field was done by Liang *et al* [60], wherein they constructed Boolean state transition tables based on the time series of gene expression and analyzed the mutual information between input and output states to infer the regulators controlling each gene. The algorithm, implemented in the REVEAL program, showed that a small number of state transition pairs are sufficient for inferring the original Boolean network correctly. Following this study, Akutsu *et al* [2] devised a simpler algorithm for the same problem and rigorously proved that if the number of incoming interactions to each node is bounded by a constant, only $O(\log n)$ state transition pairs are necessary and sufficient to correctly identify the original Boolean network with n nodes with a high probability. The algorithm was later extended to infer a noisy Boolean network since gene expression exhibits considerable uncertainty [3]. Several studies utilized probabilistic Boolean networks wherein the uncertainty in gene expression data is incorporated by using several alternative Boolean transfer functions (each with a certain probability) for a node in the network [86]. The inference process results in the Boolean transfer functions and their probabilities. Later, Laubenbacher and Stigler [55] described gene regulatory networks by time-discrete polynomial dynamical systems (an extension of Boolean modeling) and employed tools from computational algebra to design a reverse-engineering algorithm for inferring the transfer functions from experimental time-course data. More recently, Martin *et al* [62] developed a method to identify multiple gene regulatory networks that match the same time series microarray data. This method first clusters and discretizes the gene expression data using k-means clustering and support vector regression, and then samples and enumerates the Boolean networks that match the discretized data. By putting a limit on the number of inputs of a Boolean function, the methods mentioned above are able to infer gene regulatory networks efficiently.

A state-of-art Boolean network method for reverse engineering of signaling networks has been developed by Saez-Rodriguez *et al* [79]. This method, implemented in the software CellNetOptimizer, aims to bridge literature-based context-specific signaling networks, which tend to be disconnected, and protein networks inferred purely from data, which do not reflect existing mechanistic information. Their method first generates a superstructure of Boolean models which include all possible logic gates compatible with the prior literature-based signaling network. Then, this superstructure is trained against dose-response experimental data by minimizing the difference between data and model simulation while penalizing model size, and finally a set of calibrated Boolean models are obtained. Applications of this method to the growth and inflammatory signaling in human cancer cells showed that it can create Boolean network models which eliminate many false-positive interactions present in the prior literature-based signaling network and also predict new interactions [79]. Cell-specific Boolean models of immediate-early signaling networks in normal and transformed hepatocytes have been inferred by this method based on biochemical data treated with different combinations of signals and drugs, thereby revealing biochemical differences in signal transduction among normal and tumor live cells [80]. To model quantitative data in a simple and efficient way, this method was then extended to a constrained fuzzy logic-based framework which can describe intermediate levels of protein activation [64].

While Boolean networks have been widely used to infer gene regulatory relationships from time series expression profiles, few studies have so far used Boolean models to deduce regulatory mechanisms from steady-state mutant transcriptomes. Usually such data are analyzed using a simple statistical analysis that compares gene expression

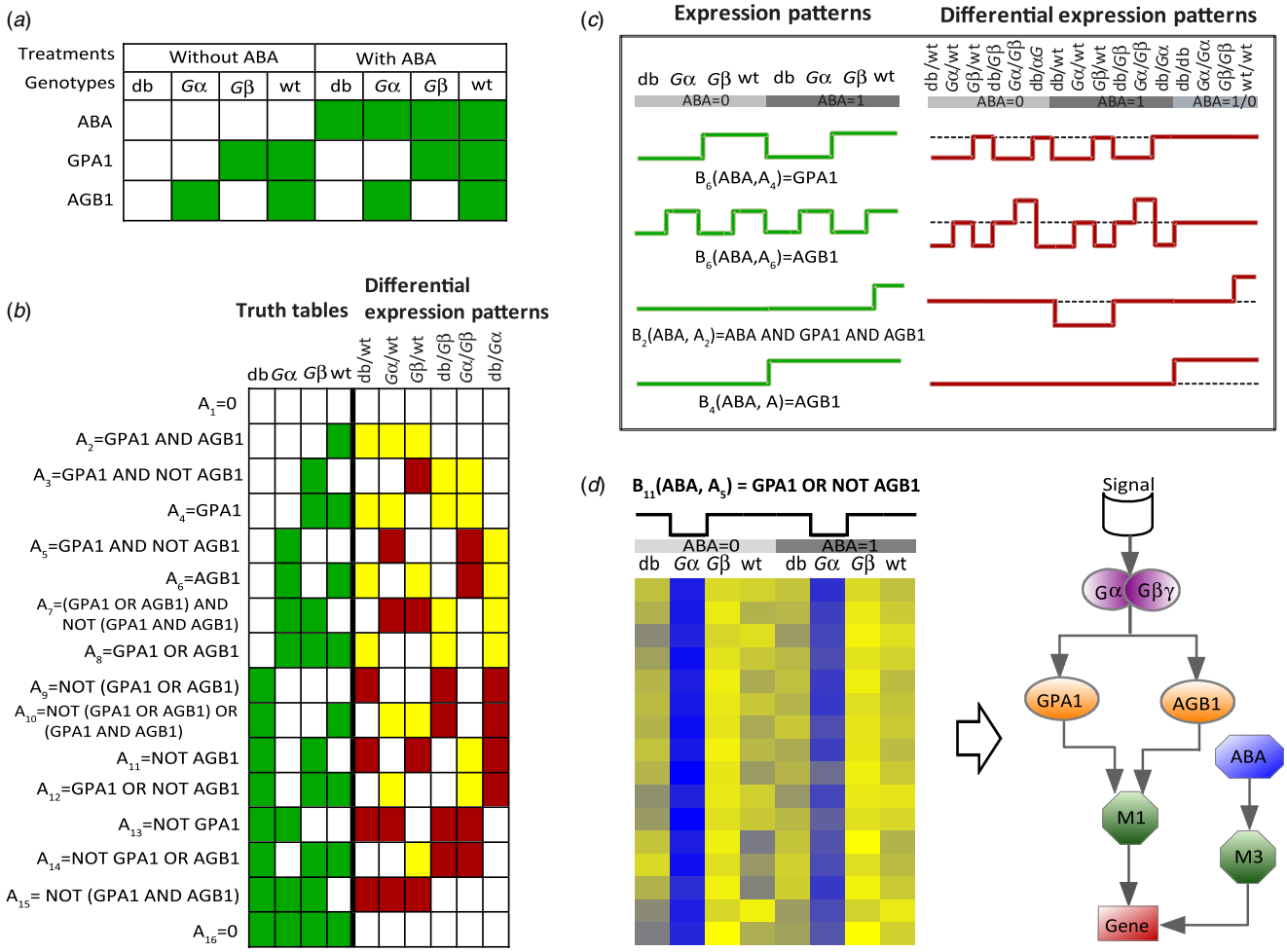

Figure 4. Application of Boolean models in analyzing complex mutant expression data. (*a*) Boolean variables ABA, GPA1, and AGB1 code the eight combinations of genotypes and treatments. Green cells represent 1, and white cells denote 0. (*b*) Boolean functions for the theoretical regulatory modes of the G-protein and the corresponding idealized differential expression patterns of genes. In the truth tables green cells represent 1, and white cells denote 0. In the differential expression patterns, red cells represent upregulation, yellow cells denote downregulation and white cells signify no differential expression. (*c*) Examples of Boolean functions for the theoretical co-regulatory modes of the G-protein and ABA, and the corresponding idealized differential expression patterns according to [71]. (*d*) An example of theoretical regulatory modes supported by enrichment of target genes. The figure is adapted from [71].

profiles in the wild type and mutants. While such an analysis can identify differentially expressed genes in each mutant, it generates results that are difficult to interpret when samples are combinatorially complex. In contrast, Boolean models are able to provide natural and intuitive interpretations about signaling and regulatory mechanisms among mutated components and signaling conditions. In an attempt to infer regulatory mechanisms of the kinase Hog1-dependent osmotic stress responses, Calpadi *et al* [21] analyzed gene expression in single- and multiple-mutant strains of budding yeast and inferred the combinatorial interactions between Hog1 and general osmotic stress (Msn2/4) pathways. In this study, simple AND and OR relations were extracted by a multiple linear regression algorithm. In [71], we developed a Boolean modeling framework to analyze transcriptomes corresponding to four genotypes and two signaling conditions. Here we briefly describe this framework to illustrate how Boolean models can be used to analyze complex steady-state mutant expression data.

Heterotrimeric G-proteins, consisting of Gα, Gβ and Gγ subunits, mediate a variety of crucial signal transduction mechanisms in both mammalian and plant systems [63, 45]. ABA is a major plant hormone which promotes tolerance of abiotic stresses such as drought and cold [51]. To answer how ABA and the G-protein interplay with each other at the transcript level, Pandey *et al* [71] generated the transcriptome data of *Arabidopsis* genome-wide genes in four genotypes (wild type, *gpa1* mutant (Gα), *agb1* mutant (Gβ) and *agb1 gpa1* double mutant (Gα, Gβ)) in the presence or absence of ABA. According to the characteristics of the sample conditions, we used Boolean variables to code the states of GPA1, AGB1 and ABA, i.e. the genotypes and treatments (figure 4(*a*)). Then the expression profile of a target gene in different genotypes and/or treatments can be seen as truth values of a Boolean function determined by the states of GPA1, AGB1 and/or ABA. The G-protein regulation was modeled by Boolean functions of the form A(GPA1, AGB1) (figure 4(*b*)) which reflect the activity of G-protein-regulated mediators that further regulate the expression level of target

genes. The co-regulation of the G-protein and ABA was described by Boolean functions of the form B(ABA, A(GPA1, AGB1)) which represent the co-regulatory activity of ABA and the G-protein-regulated mediators (see figure 4(*c*)). ABA regulation independent of the G-protein was modeled by an eight-dimensional constant vector C_{ABA}. Collectively, the Boolean modeling framework F(ABA, GPA1, AGB1) = C_{ABA} + B(ABA, A(GPA1, AGB1)) defines 142 theoretical regulatory modes of the G-protein and/or ABA which can be classified into five categories of signaling pathways [71].

After defining the Boolean functions for the regulatory modes of ABA and the G-protein, the next step is to determine which theoretically possible signaling pathways and regulatory modes are experimentally supported by the transcriptome data. To this end, genes were assigned into the theoretical regulatory modes by correlating their real differential expression patterns with the idealized differential expression patterns determined by Boolean functions (figures 4(*b*) and (*c*)). The classical Pearson correlation coefficient cannot be used for this purpose because here both shape similarity and differential expression strength are important. Thus, a new correlation measure was defined and used for associating genes with the theoretical regulatory modes [71]. Each gene was assigned to the theoretical regulatory mode for which it had the maximum correlation score (figure 4(*d*)). Then biologically realistic regulatory modes and signaling pathways were determined according to the number of associated genes (figure 4(*d*)). Overall, this Boolean modeling framework discovered new mechanisms of G-protein and hormonal control at the transcript level. The method is versatile enough to be applied to transcriptome datasets from other systems to provide new perspectives regarding switch-like signal transduction mechanisms or to infer the combinatorial regulation of transcription factors.

5. Boolean rule-based structural analysis of cellular networks

Boolean dynamic modeling is a powerful tool for analyzing the dynamic characteristics of biological systems. However, due to the exponential dependence of the size of state space on the network size, it is difficult to sufficiently sample initial conditions when the biologically relevant initial conditions are not known. Dynamic modeling also requires one to sample timescales randomly if real timescale information is not available or sufficient. These aspects make it difficult to analyze a large-scale biological network by performing Boolean dynamic simulations. Network reduction methods discussed in section 3 can alleviate these difficulties by reducing the size of the networks while preserving essential dynamic properties. Alternatively, structural methods for analysis of biological networks have been developed [53, 61, 74] as increasing evidence reveals that the structure of biological networks is closely related to their function [90, 44]. However, most of these structural methods ignore the signs of the interactions (inhibitory or activating) as well as the conditional dependence among the multiple regulators of a component, which are quite common in signal transduction and gene regulatory networks.

A notable exception is the methodology introduced by Klamt *et al* [83, 53] for structural and functional analyses of signaling and regulatory networks. This method, implemented in the software CellNetAnalyzer [52], represents a biological network by a logical interaction hypergraph whose hyper-edges connect two sets of nodes instead of two nodes. This way the relationship C* = A AND B can be represented by a hyper-edge that starts from the node set {A, B} and ends at C. Klamt *et al* used the logical steady state analysis to find minimum intervention targets, defined as a minimal set of important nodes whose simultaneous manipulation satisfies a user-defined goal (e.g. permanent deactivation of the output). Such a structural analysis can efficiently identify minimum failure modes for signaling and regulatory networks without the necessity of dynamic simulation. Recently, Abdi *et al* [1] applied fault diagnosis methods used in digital circuits to Boolean representations of signaling networks to find vulnerable components mediating a signal transduction process. The method determines the probability that the dysfunction of a signaling component propagates to the output(s) by traversing all paths from the dysfunctional site to the output(s) and by applying Boolean function-based propagation probability rules in the traversal. Abdi *et al* identified vulnerable components with high dysfunction propagation probabilities in several signal transduction networks and confirmed them by experiments [1].

In [99], we developed a structural method augmented by incorporating inhibitory regulations and conditional dependence among regulators into the topology of a signaling network. Specifically, the method introduces a complementary node for each node that has negative effects on other nodes or is inhibited by other nodes, and introduces a composite node for each set of interactions with conditional dependence. The new representation, wherein all interactions represent activation and all composite nodes indicate conditional dependence, facilitates the incorporation of conditionality in evaluating the cascading effects of node failure. The new concept of elementary signaling mode (ESM), illustrated in figure 5(*a*), was defined as the minimal set of components that are able to perform signal transduction independently [99]. We hypothesized that the signaling components whose disruption (and its cascading effects, see figure 5(*b*)) eliminates the majority or all of the ESMs are essential [99]. Validation on several signaling networks showed that this augmented structural method and essentiality criterion are in strong agreement with the results of dynamic simulations [99].

Taken together, these integrated Boolean-structural studies reveal 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 the Boolean rules and do not depend on the details of the dynamic simulation. Therefore, for exploratory analysis of large networks where dynamic modeling is computationally impractical, one can utilize Boolean rule-based structural methods to guide targeted computational or experimental design.

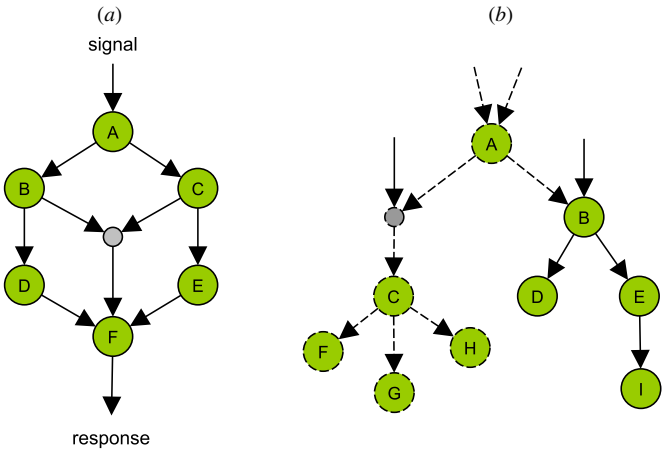

Figure 5. Illustration of ESMs in signaling networks and of the cascading effects of a component's disruption. (*a*) There are three ESMs: {A, B, D, F}, {A, B, C, F} and {A, C, E, F}. (*b*) The cascading effects of the deletion of component A. The dashed node contours and the dashed edges indicate the nodes and edges that will be disrupted in the cascading failure following the removal of node A. In (*a*) and (*b*), the small gray circles signify composite nodes.

6. Conclusions

Computational modeling of biological processes plays a pivotal role in systems biology and enables efficient *in silico* experiments whose predictions greatly improve the design of wet-bench experiments. Although Boolean network models are created with a set of assumptions and have limitations in describing the quantitative characteristics of dynamic systems, they do not require the knowledge of kinetic parameters, which make them powerful in qualitatively describing the large-scale systems' dynamics and efficiently predicting effective interventions. The success of Boolean models illustrates that in at least a subset of biological systems, the organization of the network structure plays a more important role than the kinetic details of the individual interactions [8, 58]. In addition, Boolean networks are extremely useful for modeling poorly understood large-scale systems where continuous modeling is impossible due to insufficient quantitative information. In practice, qualitative models and quantitative models complement each other. The choice between qualitative models like Boolean networks and quantitative models such as 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 for modeling regulatory and signaling networks on which more detailed continuous models can be built as kinetic information and quantitative experimental data become available [101].

Acknowledgments

This work was supported by the NSF grants CCF 0643529 (CAREER) and MCB 1121612 to RA.

References

[1] 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
[2] Akutsu T, Miyano S and Kuhara S 1999 Identification of genetic networks from a small number of gene expression patterns under the Boolean network model *Pac. Symp. Biocomput.* **4** 17–28 PMID: 10380182
[3] Akutsu T, Miyano S and Kuhara S 2000 Inferring qualitative relations in genetic networks and metabolic pathways *Bioinformatics* **16** 727–34
[4] Albert I, Thakar J, Li S, Zhang R and Albert R 2008 Boolean network simulations for life scientists *Source Code Biol. Med.* **3** 16
[5] Albert R 2007 Network inference, analysis, and modeling in systems biology *Plant Cell* **19** 3327–38
[6] Albert R, DasGupta B, Dondi R, Kachalo S, Sontag E, Zelikovsky A and Westbrooks K 2007 A novel method for signal transduction network inference from indirect experimental evidence *J. Comput. Biol.* **14** 927–49
[7] Albert R, Dasgupta B and Sontag E 2010 Inference of signal transduction networks from double causal evidence *Methods Mol. Biol.* **673** 239–51
[8] Albert R and Othmer H G 2003 The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in *Drosophila melanogaster* *J. Theor. Biol.* **223** 1–18
[9] Albert R and Wang R S 2009 Discrete dynamic modeling of cellular signaling networks *Methods Enzymol.* **467** 281–306
[10] Aldridge B B, Burke J M, Lauffenburger D A and Sorger P K 2006 Physicochemical modelling of cell signalling pathways *Nature Cell Biol.* **8** 1195–203
[11] Aldridge B B, Saez-Rodriguez J, Muhlich J L, Sorger P K and Lauffenburger D A 2009 Fuzzy logic analysis of kinase pathway crosstalk in TNF/EGF/insulin-induced signaling *PLoS Comput. Biol.* **5** e1000340
[12] Assmann S M and Albert R 2009 Discrete dynamic modeling with asynchronous update, or how to model complex systems in the absence of quantitative information *Methods Mol. Biol.* **553** 207–25
[13] Bansal M, Belcastro V, Ambesi-Impiombato A and di Bernardo D 2007 How to infer gene networks from expression profiles *Mol. Syst. Biol.* **3** 78
[14] Basso K, Margolin A A, Stolovitzky G, Klein U, Dalla-Favera R and Califano A 2005 Reverse engineering of regulatory networks in human B cells *Nature Genet.* **37** 382–90
[15] Bhalla U S and Iyengar R 1999 Emergent properties of networks of biological signaling pathways *Science* **283** 381–7
[16] Bilke S and Sjunnesson F 2002 Stability of the Kauffman model *Phys. Rev.* E **65** 016129
[17] Boole G 1848 The calculus of logic *Camb. Dublin Math. J.* **3** 183–98
[18] Bornholdt S 2008 Boolean network models of cellular regulation: prospects and limitations *J. R. Soc. Interface* **5** (Suppl. 1) S85–94
[19] Brown P O and Botstein D 1999 Exploring the new world of the genome with DNA microarrays *Nature Genet.* **21** 33–7
[20] Campbell C, Yang S, Albert R and Shea K 2011 A network model for plant-pollinator community assembly *Proc. Natl Acad. Sci. USA* **108** 197–202
[21] Capaldi A P, Kaplan T, Liu Y, Habib N, Regev A, Friedman N and O'Shea E K 2008 Structure and function of a transcriptional network activated by the MAPK Hog1 *Nature Genet.* **40** 1300–6

[22] Chaouiya C 2007 Petri net modelling of biological networks *Brief Bioinform.* **8** 210–9

[23] Chaves M, Albert R and Sontag E D 2005 Robustness and fragility of Boolean models for genetic regulatory networks *J. Theor. Biol.* **235** 431–49

[24] Chaves M, Sontag E D and Albert R 2006 Methods of robustness analysis for Boolean models of gene control networks *Syst. Biol.* **153** 154–67

[25] Chen L, Wang R, Li C and Aihara K 2010 *Modelling Biomolecular Networks in Cells: Structures and Dynamics* (London: Springer)

[26] de Jong H, Geiselmann J, Hernandez C and Page M 2003 Genetic Network Analyzer: qualitative simulation of genetic regulatory networks *Bioinformatics* **19** 336–44

[27] de Jong H, Gouze J L, Hernandez C, Page M, Sari T and Geiselmann J 2004 Qualitative simulation of genetic regulatory networks using piecewise-linear models *Bull. Math. Biol.* **66** 301–40

[28] Di Cara A, Garg A, De Micheli G, Xenarios I and Mendoza L 2007 Dynamic simulation of regulatory networks using SQUAD *BMC Bioinform.* **8** 462

[29] Dubrova E and Teslenko M 2011 A SAT-based algorithm for finding attractors in synchronous Boolean networks *IEEE/ACM Trans. Comput. Biol. Bioinform.* **8** 1393–9

[30] Farrow C, Heidel J, Maloney J and Rogers J 2004 Scalar equations for synchronous Boolean networks with biological applications *IEEE Trans. Neural Netw.* **15** 348–54

[31] Faure 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** e124–31

[32] Friedman D B 2007 Quantitative proteomics for two-dimensional gels using difference gel electrophoresis *Methods Mol. Biol.* **367** 219–39 PMID: 17185779

[33] Garg A, Di Cara A, Xenarios I, Mendoza L and De Micheli G 2008 Synchronous versus asynchronous modeling of gene regulatory networks *Bioinformatics* **24** 1917–25

[34] Giacomantonio C E and Goodhill G J 2010 A Boolean model of the gene regulatory network underlying mammalian cortical area development *PLoS Comput. Biol.* **6** e1000936

[35] Glass L 1975 Classification of biological networks by their qualitative dynamics *J. Theor. Biol.* **54** 85–107

[36] 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

[37] Grinstead C and Snell J 1997 *Introduction to Probability* (Providence, RI: American Mathematical Society)

[38] Gross L 2006 When evidence is scant, mathematical modeling offers a roadmap for discovery *PLoS Biol.* **4** e323

[39] Hallikas O and Taipale J 2006 High-throughput assay for determining specificity and affinity of protein-DNA binding interactions *Nature Protoc.* **1** 215–22

[40] Harvey I and Bossomaier T 1997 Time out of joint: attractors in asynchronous random Boolean networks *Proc. of the 4th European Conf. on Artificial Life (ECAL97)* ed P Husbands and I Harvey (Cambridge, MA: MIT Press) pp 67–75

[41] Heidel J, Maloney J, Farrow C and Rogers J 2003 Finding cycles in synchronous Boolean networks with applications to biochemical systems *Int. J. Bifurcation Chaos* **13** 535–52

[42] 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

[43] Hinkelmann F, Brandon M, Guang B, McNeill R, Blekherman G, Veliz-Cuba A and Laubenbacher R 2011 ADAM: analysis of discrete models of biological systems using computer algebra *BMC Bioinform.* **12** 295

[44] Jeong H, Mason S P, Barabasi A L and Oltvai Z N 2001 Lethality and centrality in protein networks *Nature* **411** 41–2

[45] Jones A M and Assmann S M 2004 Plants: the latest model system for G-protein research *EMBO Rep.* **5** 572–8

[46] Kachalo S, Zhang R, Sontag E, Albert R and DasGupta B 2008 NET-SYNTHESIS: a software for synthesis, inference and simplification of signal transduction networks *Bioinformatics* **24** 293–5

[47] Kahlem P and Birney E 2006 Dry work in a wet world: computation in systems biology *Mol. Syst. Biol.* **2** 40

[48] Karlebach G and Shamir R 2008 Modelling and analysis of gene regulatory networks *Nature Rev. Mol. Cell Biol.* **9** 770–80

[49] Kauffman S A 1969 Metabolic stability and epigenesis in randomly constructed genetic nets *J. Theor. Biol.* **22** 437–67

[50] Kim J K, Bamba T, Harada K, Fukusaki E and Kobayashi A 2007 Time-course metabolic profiling in *Arabidopsis thaliana* cell cultures after salt stress treatment *J. Exp. Bot.* **58** 415–24

[51] Kim T H, Bohmer M, Hu H, Nishimura N and Schroeder J I 2010 Guard cell signal transduction network: advances in understanding abscisic acid, CO_2, and Ca^{2+} signaling *Annu. Rev. Plant Biol.* **61** 561–91

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

[53] Klamt S, Saez-Rodriguez J, Lindquist J A, Simeoni L and Gilles E D 2006 A methodology for the structural and functional analysis of signaling and regulatory networks *BMC Bioinform.* **7** 56

[54] Kusano M *et al* 2011 Metabolomics reveals comprehensive reprogramming involving two independent metabolic responses of Arabidopsis to UV-B light *Plant J.* **67** 354–69

[55] Laubenbacher R and Stigler B 2004 A computational algebra approach to the reverse engineering of gene regulatory networks *J. Theor. Biol.* **229** 523–37

[56] Lee T I *et al* 2002 Transcriptional regulatory networks in *Saccharomyces cerevisiae Science* **298** 799–804

[57] Lee W P and Tzou W S 2009 Computational methods for discovering gene networks from expression data *Brief Bioinform.* **10** 408–23

[58] Li F, Long T, Lu Y, Ouyang Q and Tang C 2004 The yeast cell-cycle network is robustly designed *Proc. Natl Acad. Sci. USA* **101** 4781–6

[59] 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** e312

[60] Liang S, Fuhrman S and Somogyi R 1998 Reveal, a general reverse engineering algorithm for inference of genetic network architectures *Pac. Symp. Biocomput.* **3** 18–29 PMID: 9697168

[61] Liu W, Li D, Zhang J, Zhu Y and He F 2006 SigFlux: a novel network feature to evaluate the importance of proteins in signal transduction networks *BMC Bioinform.* **7** 515

[62] Martin S, Zhang Z, Martino A and Faulon J L 2007 Boolean dynamics of genetic regulatory networks inferred from microarray time series data *Bioinformatics* **23** 866–74

[63] Morris A J and Malbon C C 1999 Physiological regulation of G protein-linked signaling *Physiol. Rev.* **79** 1373–430 PMID: 10508237

[64] Morris M K, Saez-Rodriguez J, Clarke D C, Sorger P K and Lauffenburger D A 2011 Training signaling pathway maps to biochemical data with constrained fuzzy logic:

quantitative analysis of liver cell responses to inflammatory stimuli *PLoS Comput. Biol.* **7** e1001099

[65] Morris M K, Saez-Rodriguez J, Sorger P K and Lauffenburger D A 2010 Logic-based models for the analysis of cell signaling networks *Biochemistry* **49** 3216–24

[66] Mortazavi A, Williams B A, McCue K, Schaeffer L and Wold B 2008 Mapping and quantifying mammalian transcriptomes by RNA-Seq *Nature Methods* **5** 621–8

[67] Mussel C, Hopfensitz M and Kestler H A 2010 BoolNet—an R package for generation, reconstruction and analysis of Boolean networks *Bioinformatics* **26** 1378–80

[68] Naldi A, Remy E, Thieffry D and Chaouiya C 2011 Dynamically consistent reduction of logical regulatory graphs *Theor. Comput. Sci.* **412** 2207–18

[69] Olsen J V, Blagoev B, Gnad F, Macek B, Kumar C, Mortensen P and Mann M 2006 Global, *in vivo*, and site-specific phosphorylation dynamics in signaling networks *Cell* **127** 635–48

[70] Ong S E, Foster L J and Mann M 2003 Mass spectrometric-based approaches in quantitative proteomics *Methods* **29** 124–30

[71] Pandey S, Wang R S, Wilson L, Li S, Zhao Z, Gookin T E, Assmann S M and Albert R 2010 Boolean modeling of transcriptome data reveals novel modes of heterotrimeric G-protein action *Mol. Syst. Biol.* **6** 372

[72] Papin J A, Hunter T, Palsson B O and Subramaniam S 2005 Reconstruction of cellular signalling networks and analysis of their properties *Nature Rev. Mol. Cell Biol.* **6** 99–111

[73] Peng S C, Wong D S, Tung K C, Chen Y Y, Chao C C, Peng C H, Chuang Y J and Tang C Y 2010 Computational modeling with forward and reverse engineering links signaling network and genomic regulatory responses: NF-κB signaling-induced gene expression responses in inflammation *BMC Bioinform.* **11** 308

[74] Potapov A P, Goemann B and Wingender E 2008 The pairwise disconnectivity index as a new metric for the topological analysis of regulatory networks *BMC Bioinform.* **9** 227

[75] Richardson K A 2005 Simplifying Boolean networks *Adv. Complex Syst.* **8** 365–81

[76] Ruths D, Muller M, Tseng J T, Nakhleh L and Ram P T 2008 The signaling petri net-based simulator: a non-parametric strategy for characterizing the dynamics of cell-specific signaling networks *PLoS Comput. Biol.* **4** e1000005

[77] Saadatpour A, Albert I and Albert R 2010 Attractor analysis of asynchronous Boolean models of signal transduction networks *J. Theor. Biol.* **266** 641–56

[78] Saadatpour A, Wang R S, Liao A, Liu X, Loughran T P, Albert I and Albert R 2011 Dynamical and structural analysis of a T cell survival network identifies novel candidate therapeutic targets for large granular lymphocyte leukemia *PLoS Comput. Biol.* **7** e1002267

[79] Saez-Rodriguez J, Alexopoulos L G, Epperlein J, Samaga R, Lauffenburger D A, Klamt S and Sorger P K 2009 Discrete logic modelling as a means to link protein signalling networks with functional analysis of mammalian signal transduction *Mol. Syst. Biol.* **5** 331

[80] Saez-Rodriguez J, Alexopoulos L G, Zhang M, Morris M K, Lauffenburger D A and Sorger P K 2011 Comparing signaling networks between normal and transformed hepatocytes using discrete logical models *Cancer Res.* **71** 5400–11

[81] Saez-Rodriguez J *et al* 2007 A logical model provides insights into T cell receptor signaling *PLoS Comput. Biol.* **3** e163

[82] 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

[83] Samaga R, Von Kamp A and Klamt S 2010 Computing combinatorial intervention strategies and failure modes in signaling networks *J. Comput. Biol.* **17** 39–53

[84] Schlatter R, Philippi N, Wangorsch G, Pick R, Sawodny O, Borner C, Timmer J, Ederer M and Dandekar T 2012 Integration of Boolean models exemplified on hepatocyte signal transduction *Brief Bioinform.* **13** 365–76 PMID: 22016404

[85] Schlatter R, Schmich K, Avalos Vizcarra I, 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** e1000595

[86] 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** 261–74

[87] Skodawessely T and Klemm K 2011 Finding attractors in asynchronous Boolean dynamics *Adv. Complex Syst.* **14** 439–49

[88] Sobie E A, Lee Y S, Jenkins S L and Iyengar R 2011 Systems biology—biomedical modeling *Sci. Signal.* **4** tr2

[89] Srividhya J, Crampin E J, McSharry P E and Schnell S 2007 Reconstructing biochemical pathways from time course data *Proteomics* **7** 828–38

[90] Stelling J, Klamt S, Bettenbrock K, Schuster S and Gilles E D 2002 Metabolic network structure determines key aspects of functionality and regulation *Nature* **420** 190–3

[91] Thakar J and Albert R 2010 Boolean models of within-host immune interactions *Curr. Opin. Microbiol.* **13** 377–81

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

[93] Thakar J, Saadatpour-Moghaddam A, Harvill E T and Albert R 2009 Constraint-based network model of pathogen-immune system interactions *J. R. Soc. Interface* **6** 599–612

[94] Thomas R 1973 Boolean formalization of genetic control circuits *J. Theor. Biol.* **42** 563–85

[95] Tong A H *et al* 2004 Global mapping of the yeast genetic interaction network *Science* **303** 808–13

[96] Veliz-Cuba A 2011 Reduction of Boolean network models *J. Theor. Biol.* **289** 167–72

[97] von Dassow G, Meir E, Munro E M and Odell G M 2000 The segment polarity network is a robust developmental module *Nature* **406** 188–92

[98] Walsh E R, Thakar J, Stokes K, Huang F, Albert R and August A 2011 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.* **186** 2936–49

[99] Wang R S and Albert R 2011 Elementary signaling modes predict the essentiality of signal transduction network components *BMC Syst. Biol.* **5** 44

[100] Watterson S, Marshall S and Ghazal P 2008 Logic models of pathway biology *Drug. Discov. Today* **13** 447–56

[101] Wittmann D M, Krumsiek J, Saez-Rodriguez J, Lauffenburger D A, Klamt S and Theis F J 2009 Transforming Boolean models to continuous models: methodology and application to T-cell receptor signaling *BMC Syst. Biol.* **3** 98

[102] Yu H *et al* 2008 High-quality binary protein interaction map of the yeast interactome network *Science* **322** 104–10

[103] Zhang R, Shah M V, Yang J, Nyland S B, Liu X, Yun J K, Albert R and Loughran T P Jr 2008 Network model of survival signaling in large granular lymphocyte leukemia *Proc. Natl Acad. Sci. USA* **105** 16308–13

[104] Zhao Q 2005 A remark on 'scalar equations for synchronous Boolean networks with biological applications' by C Farrow, J Heidel, J Maloney, and J Rogers *IEEE Trans. Neural Netw.* **16** 1715–6

[105] Zhao Z, Zhang W, Stanley B A and Assmann S M 2008 Functional proteomics of *Arabidopsis thaliana* guard cells uncovers new stomatal signaling pathways *Plant Cell* **20** 3210–26

[106] Zheng J, Zhang D, Przytycki P F, Zielinski R, Capala J and Przytycka T M 2010 SimBoolNet—a Cytoscape plugin for dynamic simulation of signaling networks *Bioinformatics* **26** 141–2