Cell Signaling Dynamics Analysis in Leukemia with Switching Boolean Networks

Woochang Hwang[1]   Doheon Lee[1]

1 Department of Bio and Brain Engineering, KAIST, Republic of Korea

Abstract A major challenge in systems biology is to make and analyze models of signaling networks. Here, we suggest the use of switching Boolean networks using threshold Boolean networks; we made a model of the acute myeloid leukemia (AML) signaling network and analyzed this network to find the component that makes the signaling network abnormal by being deregulated. Acute myeloid leukemia (AML) is characterized by the rapid growth of abnormal white blood cells, which accumulate in the patient's bone marrow and perturb the production of normal blood cells. We constructed a model of the AML signaling network by combining the signaling pathways involved in either myeloid differentiation or cell proliferation. We then analyzed this signaling network using switching Boolean networks. Some of the components found in by this simulation had been previously experimentally validated by other researchers; however, we also discovered some new components through our technique.

Keywords AML; Acute Myeloid Leukemia; Signaling Network; Boolean Networks;

1   Introduction

Developing technologies such as DNA microarrays, high-throughput genome sequencing and other high throughput technologies have led to a rapid increase in the amount of biological experiment data that is available. The information generated by these methods exceeds the human capacity for analysis and requires computational power to integrate information in high throughput data. Due to the large amounts of biological data available, researchers have started to combine the data in order to understand biology at the systems level. Systems biology aims to understand biological processes and cellular functions through the modeling and simulation of biological systems [1].

There are many formalisms for modeling of biological systems [2]. Of these, qualitative modeling has many advantages. The advantages are that it only needs topological information such as interactions between substrates of model systems, and it can be used even with incomplete knowledge of a system.

Qualitative modeling has successfully modeled many different biological systems. For instance, one study robustly modeled the yeast cell cycle using

threshold Boolean networks with a synchronous update [3]. However, they did not examine differences in the speed of signal propagation. Another study found a key component in network modeling of survival signaling in large granular lymphocyte leukemia [4]. They constructed a T-LGL signal network and used a Boolean model for the networks' dynamics. They reproduced the signaling abnormalities and predicted a key mediator in the signaling network. However, they did not represent protein self-degradation. In both studies, Boolean networks were used to model biological systems.

Boolean networks are widely used to model regulatory and signaling networks because of their straightforwardness, robustness and compatibility with qualitative data. However, it is hard to specify the output of all of the combinations of input nodes because of limited information. To overcome this limitation of Boolean networks, threshold Boolean networks are made. Threshold Boolean networks are a subset of Boolean networks that have a Boolean function for each node that depends only on the sum of its input signals.

Threshold Boolean networks also have limitations. This modeling technique relies completely on network information. Therefore, if a network is incomplete there will be gaps between the biological reality and the model's output, resulting in a modeling anomaly. Another limitation is the synchronous update algorithm. It cannot measure differences in the speed of signal propagation. In biological systems, no two cells have exactly the same properties, resulting in differences in signal propagation between cells. Therefore, synchronous update algorithms are not proper for simulating biological systems.

In this study, we make a simulation model which is called switching Boolean networks. This model overcomes the limitations of threshold Boolean networks, which are model anomalies and the use of the synchronous update algorithm. We found the essential components in an AML signal network with it and suggest it as an improved modeling technique for biological data.

2   Switching Boolean Networks

To overcome modeling anomaly and synchronous update problem, we modified threshold Boolean networks.

2.1 Switching Boolean Networks

In Biology, interactions between proteins occur at different times and a dynamic model would involve various binding constants and binding rates. However, biological signaling networks seemed to be reflected by the on/off characteristics of nodes in a network. Here, we use a simplified dynamic network, as we want to see a whole picture of the signaling network of biology rather than an extremely detailed view.

Network nodes represent proteins or cellular functions, such as differentiation, proliferation and apoptosis. Each node in a signaling network has only two states, $S_i = 1$ or $S_i = 0$, which represent an active or inactive state of the node, respectively. The biological meaning is that the protein is either active or not due to different biochemical mechanisms such as gene expression or post-translational modifications

such as phosphorylation, which determine if a protein is active. The active state means that a small molecule is produced, a transcript is produced and translated, or a protein or cellular function is activated. The inactive state indicates the absence of a small molecule or transcript or that a protein or cellular function is inhibited.[4] The state of the node in the next time step is determined by the states of source nodes in the present time step by the application of the following rule:

$$S_i(t+1) = \begin{cases} 1, \sum_j a_{ij}S_j(t) > \theta_i \\ 0, \sum_j a_{ij}S_j(t) < \theta_i \\ S_i(t), \sum_j a_{ij}S_j(t) = \theta_i \end{cases} \quad (1)$$

where $a_{ij}(t) = a_a$ for an activator edge from node j to node i and $a_{ij}(t) = a_h$ for an inhibitor edge from node j to node i.[3] In principal, when an inhibitor protein for the target protein is active, the state of the target protein becomes inactive, even if the activator protein for the same target is active [5]. To meet this biological principal, we assigned a weight of 1 to the protein activator edge and the weight of the protein inhibitor edge is $-\left(\sum_j a_{aj} + 1\right)$. However, the action of a cellular function node is different than that of a protein node in a signaling network. There is no physical interaction between a protein and its cellular function. To model cellular functions with switching Boolean networks, we applied different edge weights for source nodes of a cellular function node (e.g., proliferation, differentiation and apoptosis) according to the protein's function in these nodes. θ_i is the activation threshold of node i, which is set to 0 for all nodes. Because of the incomplete data for the biological network, there are nodes that have no activator and inhibitor nodes. To overcome these model anomalies, we inserted an external activator or inhibitor node for nodes that have no activator or inhibitor node, respectively. The state of an external node is randomly assigned to model differences between cells. We use a random-order asynchronous updating algorithm to model differences in the speed of signal propagation in cells. Multiple replicated simulations are performed. Degradation of a protein is modeled by turning the state of the protein to off randomly.

2.2 Acute Myeloid Leukemia

Acute Myeloid Leukemia (AML) is characterized by the rapid growth of abnormal white blood cells, which are granulocyte or monocyte precursors that accumulate in the bone marrow and blood. These cells interfere with the production of normal blood cells by causing problems with differentiation and proliferation [6]. There are three possible genetic mechanisms that could cause the block in differentiation [6]. One possibility is that a disruption of cell-cycle control could block differentiation. Another possibility is secondary events in carcinogenesis. The last one is that the disruption of certain gene products affects both the cell cycle and the differentiation of the blood cells. Recent evidence suggests that the third proposed genetic mechanism is important for AML. Therefore, it is important to

find the gene products that can effect proliferation or the differentiation of these cells in order to overcome AML. Here is our strategy for identifying the essential components of AML. First, we made an AML signaling network using data extracted from the literature. Second, we simulated the AML signaling network using switching Boolean networks.

Fourth, we identified possible causes of network abnormalities. Finally, we identified the essential components of AML. To make the AML signaling network, we first constructed a general granulopoiesis network from the literature. We regarded a normal granulopoiesis network as a framework for an AML signaling network. We will refer to this network as the "original network". The Original network has a source (upstream regulator) node, a target (downstream regulator) node and interactions between the source node and the target node. The biological meaning of these interactions is to "promote", "activate" or "inhibit". To understand the effect of altered granulopoiesis conditions on normal cell differentiation and proliferation, we augmented the nodes in the original network, which is known deregulated in AML. To simplify the original network, we did two things. First, we removed unconnected nodes in the original network, which will not affect the simulation results. Second, if there are two relationships such that A inhibits C, A inhibits B, and B activates C, we removed the B node in the original network. After constructing the AML signaling network, we simulated this network using switching Boolean networks.

2.3 Simulation

In the biological system, there is a different time scale between the state change from the regulators and the state change of the regulators' targets. The enzyme propagation time is rarely known from experiments. Thus, we used an asynchronous updating algorithm that simulates the different signal propagation speeds [4]. To equally simulate all possible timescales, a random-order asynchronous algorithm was used. In this algorithm, the time step is a round of updating, during which all nodes are updated in a randomly selected order [4]. The updating scheme of an asynchronous algorithm is written as rule (1). To reproduce how a group of cells responds to the same initial signal and to simulate variability among cells, we performed multiple simulations with same initial conditions but different updating orders. The state of the receptors (G-CSF, GM-CSF and TGFβ) was set to active at the beginning of every simulation. The frequency of the activate state of a node during a simulation is quantified by:

$$F_i^t = \sum_{j=1}^{N} \frac{S_i^t(j)}{N} \quad (2)$$

F_i^t stands for the frequency of the active state of node i at time step t; N is total number of simulations, and $S_i^t(j)$ is the status of node i at time step t in the jth simulation.

To reveal possible causes of network abnormalities, we turned on or off every node in each simulation.

Figure 1. Algorithm of Switching Boolean Networks

Algorithm: Switching Boolean Networks
Input : Network
Output: The frequency of node activation

$$F_i(t) = \frac{\sum_{n=1}^{N} S_i^n(t)}{N}$$

$S_i(0)$ = random $(0,1)$ for $\underline{i}$ = node in Network
For $n = 1$ to N = number of simulations $(200)\{$
    For $t = 1$ to T = number of time steps $(50)\{$
    For $p = 1$ to P = number of nodes$\{$
      $\underline{i}$ = randomly selected node
      $\Omega_{\underline{i}}$ = an index set of input nodes of $\underline{i}$

$$S_i(t) = \begin{cases} 0, & \sum_{j\in\Omega_i} a_{ji} S_j(t) < \theta_i \\ 1, & \sum_{j\in\Omega_i} a_{ji} S_j(t) > \theta_i \\ S_i(t), & \sum_{j\in\Omega_i} a_{ji} S_j(t) = \theta_i \end{cases}$$

    $\}$
  $\}$
$\}$
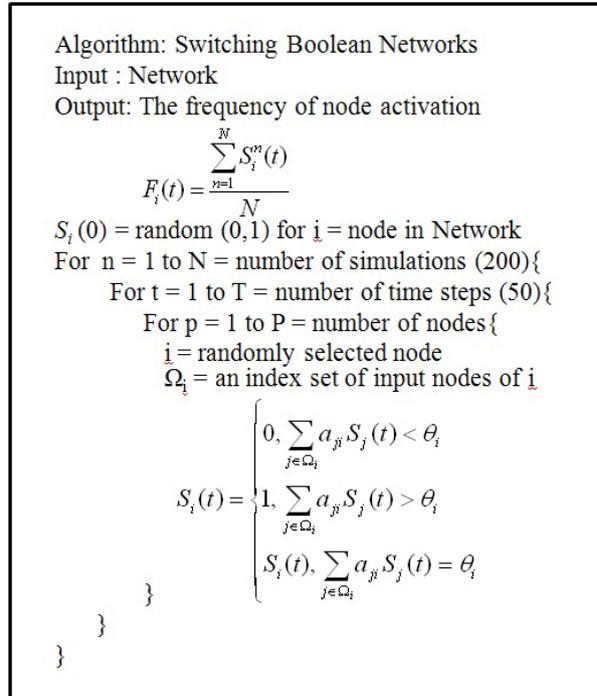
Figure 2. AML signaling network with 61 nodes and 107 edges.
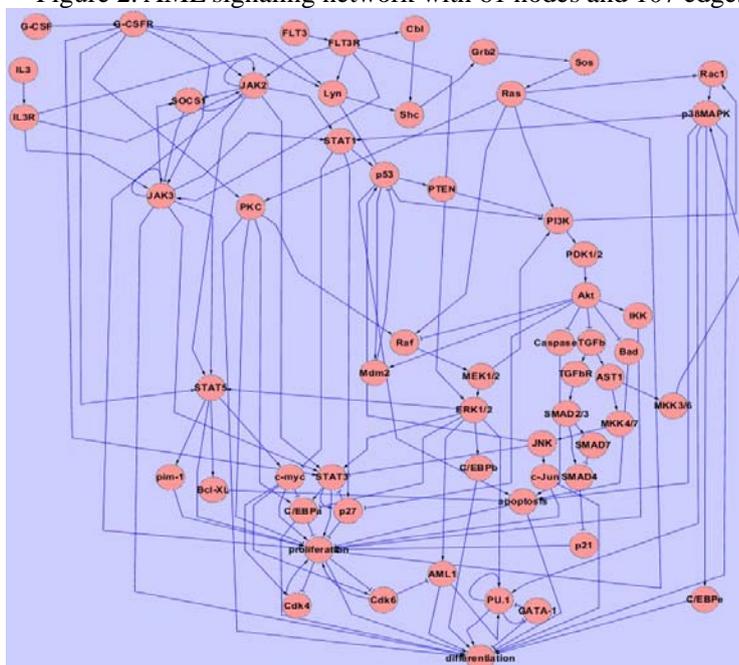
3   Results and discussion

3.1 Performance comparison

To verify the simulation performance of switching Boolean networks, we compared it to synchronous threshold Boolean networks. In this test, we turned off known drug targets of AML such as C/EBPα, AML1 and PU.1.

Figure 3 a) Synchronous threshold Boolean networks b) Switching Boolean networks
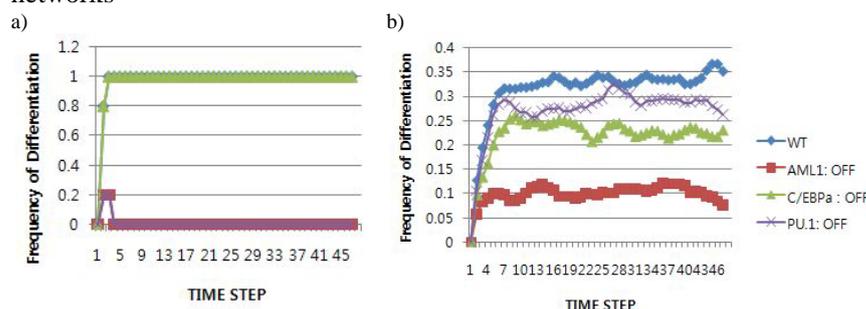
3.2 Switching a single node off

Table 1 Simulation results from switching a single node off in an AML signaling network.

Nodes that down-regulate differentiation		Nodes that affect up-regulate differentiation	
Validated proteins		Validated proteins	
AML1	STAT1	PTEN	P21
ERK1/2	c-Jun	P27	C/EBPα
P27	Bcl-XL	Candidate proteins	
JAK3	C/EBPε	SMAD7	PKC
C/EBPα	SMAD7	STAT1	P38MAPK
PU.1	PTEN		
p21	p53		
p38MAPK	Rac1		
Candidate proteins			
PKC	MKK4/7		
TGFb	AST1		
G-CSF	STAT5		
JNK			

In the case of nodes that down-regulate differentiation by being turned off, AML1-ETO down-regulates granulocyte differentiation.[7] STAT1 inhibits monocyte differentiation.[8] An example of a node that promotes proliferation by being turned off is C/EBPα. The loss of C/EBPα cell cycle control increases myeloid progenitor proliferation.[9] These results also give us candidate proteins that could make inhibit cell differentiation or promote proliferation.

3.3 Switching a single node on

In the case of nodes that down-regulate differentiation by being switched on, the overexpression of PDK1 was found to be a common feature of acute myeloid leukemia [10]. In the case of nodes that promote proliferation by being switched on, overexpression of c-Jun in AML cells was able to induce proliferation [11]. These results also give us candidate proteins that can inhibit cell differentiation or induce proliferation.

Table 2 Simulation results from switching a single node on in the AML signaling network.

Nodes that down-regulate differentiation		Nodes that affect up-regulate differentiation	
Validated proteins		Validated proteins	
GATA-1	c-Myc	c-Myc	Ras
Cdk6	Akt	STAT5	Akt
SOCS1	STAT5	Cdk6	c-Jun
JAK2	PDK1/2	JAK2	Cdk4
Pim-1	Ras	IKK	FLT3
IKK		Pim-1	PI3K
Candidate proteins		STAT3	ERK1/2
P27	SMAD2/3	Mdm2	Raf
SMAD4	Mdm2	Candidate proteins	
		SMAD2/3	SMAD4
		Bcl-XL	PDK1/2
		Sos	Grb2

4   Conclusion

Boolean networks are one type of discrete dynamic modeling method. It is straightforward, robust and compatible with qualitative data. Our proposed method can overcome modeling anomalies and differences in the speed of signal propagation in the network.

We simulated an AML signal network to find essential components that could cause AML using switching Boolean networks. Our results found that, for the simulation that switched a single node off, 75% of the proteins which is affected the differentiation block and 55% of the proteins which is affected hyperproliferation are found at literatures; for the simulation that switched a single node on, 75% of the proteins which is affected the differentiation block and 54% of the proteins which is affected hyperproliferation are found at literatures.

We were able to predict essential components that could be key mediators of the AML signal network.

We suspect that some of these predicted essential components could be candidate targets for AML drugs.

Acknowledgements

References

[1]   H. Kitano. Systems Biology: A Brief Overview. 2002, 295: 1662-1664.

[2]   H. de Jong. Modeling and Simulation of Genetic Regulatory Systems: A Literature Review. Journal of Computational Biology,2002, 9: 67-103.

[3]   F. Li, T. Long, Y. Lu, Q. Ouyang, and C. Tang. The yeast cell-cycle network is robustly designed.PNAS, 2004, 101: 4781-4786.

[4]   R. Zhang, M. V. Shah, J. Yang, S. B. Nyland, X. Liu, J. K. Yun, R. Albert, and T. P. Loughran. Network model of survival signaling in large granular lymphocyte leukemia.PNAS ,2008, 105: 16308-16313.

[5]   B. N. Kholodenko. Cell-signalling dynamics in time and space. Nat Rev Mol Cell Biol,2006, 7: 165-176.

[6]   D. G. Tenen. Disruption of differentiation in human cancer: AML shows the way. Nat Rev Cancer,2003, 3: 89-101.

[7]   T. Pabst, B. U. Mueller, N. Harakawa, C. Schoch, T. Haferlach, G. Behre, W. Hiddemann, D.-E. Zhang, and D. G. Tenen. AML1-ETO downregulates the granulocytic differentiation factor C/EBPα in t(8;21) myeloid leukemia. Nat Med,2001, 7: 444-451.

[8]   A. Dimberg, K. K. ehed, K. Nilsson, and F. erg. Inhibition of Monocytic Differentiation by Phosphorylation-deficient Stat1 is Associated with Impaired Expression of Stat2, ICSBP/IRF8 and C/EBPε. Scandinavian Journal of Immunology,2006, 64: 271-279.

[9]   B. T. Porse, D. Bryder, K. Theilgaard-Monch, M. S. Hasemann, K. Anderson, I. Damgaard, S. E. W. Jacobsen, and C. Nerlov. Loss of C/EBPα cell cycle control increases myeloid progenitor proliferation and transforms the neutrophil granulocyte lineage. JEM,2005, 202: 85-96.

[10] L. Pearn, J. Fisher, A. K. Burnett, and R. L. Darley. The role of PKC and PDK1 in monocyte lineage specification by Ras. Blood,2007, 109: 4461-4469.

[11] J. Rangatia, R. K. Vangala, S. M. Singh, A. A. Peer Zada, A. Elsasser, A. Kohlmann, T. Haferlach, D. G. Tenen, W. Hiddemann, and G. Behre. Elevated c-Jun expression in acute myeloid leukemias inhibits C/EBPα DNA binding via leucine zipper domain interaction. Oncogene,2003, 22: 4760-4764.