31st Annual International Conference of the IEEE EMBS
Minneapolis, Minnesota, USA, September 2-6, 2009

Identification of Critical Molecules via Fault Diagnosis Engineering

Ali Abdi[1], Mehdi B. Tahoori[2], Effat S. Emamian[3]

[1]New Jersey Institute of Technology, [2]Northeastern University, [3]Advanced Technologies for Novel Therapeutics (ATNT)

Abstract — **Systems biology envisions that the application of complex system engineering approaches to cell signaling molecular networks can lead to novel understandings of complex human disorders. In this paper we show that by developing biologically-driven vulnerability assessment methods, the vulnerability of complex signaling networks to the dysfunction of each molecule can be determined. We have analyzed signaling networks that regulate mitosis and the activity of the transcription factor CREB. Our results indicate that biologically-relevant critical components of intracellular molecular networks can be identified using the proposed systems biology/fault diagnosis engineering technique. The application of this approach can improve our physiological understanding of the functionality of biological systems, can be used as a tool to identify novel genes associated with complex human disorders, and ultimately, has the potential to find the most prominent targets for drug discovery.**

I. Introduction

Molecular biologists have collected enormous amount of information with regard to the function of individual signaling molecules. Now we know many detailed individual molecular mechanisms which regulate the cellular function. In recent years, systems biologists have started to integrate these individual interactions and components, to analyze the properties and functions that emerge from these complex biological systems. Targeting the activity of different enzymes has been a strategy for the development of novel therapeutics. However, curative treatments for some of the most common human diseases such as cancer, neurodegenerative and psychiatric brain disorders are still unknown. This is because the pathology of such complex disorders results from the failure of a highly sophisticated biological *system* and not just a single molecule or a specific pathway. So, to understand the pathology and to move towards the effective treatments, a *systems biology* approach needs be developed, to study the orchestrated function of complex molecular pathways.

A cell in the human body includes many biomolecules which interact with each other through a network of cell signaling pathways. Dysfunction of each molecule negatively affects the function of signaling networks. This in turn will ultimately result in a transition from the normal function (physiological condition) to a dysfunctional system (diseased or pathological condition) [1].

For some of the hereditary human diseases such as Huntington's disease, variation in a single gene is known to be the cause for the pathology. However, there are numerous prevalent and complex human disorders such as cancer, neurodegenerative and psychiatric brain disorders and metabolic disorders, that are not caused by the dysfunction of a single gene. In these complex trait disorders, it is not known which molecules, and to what extent, may have causative effects.

The main theme of this paper is to introduce and develop a new engineering-oriented systems biology approach, where the disease is considered as a *faulty complex system*. The failure of this complex system, which is composed of numerous interconnected pathways, is due to the dysfunction of one or several molecules that fail to respond correctly to the upstream signals. The ultimate goal of this approach is to answer this question in molecular cell biology: "Which components are the critical players in the development of the pathology (failure of the system)?". To answer this question, we propose a new approach, inspired by circuit fault diagnosis engineering concepts [2]. This approach is specially designed to analyze complex intracellular signaling networks. Equipped with this bioengineering approach, the goal is to identify those key signaling molecules which have causative effects in complex human disorders. These key molecules are appropriate targets for drug development.

II. Molecular Networks , Circuits, Fault Diagnosis

It is well recognized that there are many similarities between digital electronic circuits and genetic/signaling networks [3]. In a manufacturing facility, a digital circuit is manufactured based on a particular design and is supposed to provide a specific function. However, during the fabrication process, physical defects such as faulty transistors, open and short wires, etc., may happen, which cause the manufactured circuit not to function correctly (according to the design specifications) [2]. Testing of digital circuits and systems is therefore necessary to separate defective manufactured parts from the non-defective ones, in order to guarantee the delivery of fault-free products to the customers. The test itself is an assessment of the manufactured circuit, according to a set of criteria [2]. During the lifetime operation of electronic systems, the correct functionality is a key aspect and is typically referred to as reliability.

In order to determine the most critical molecules in a network of molecules, we propose to take advantage of a

class of electronic circuit reliability analysis techniques known as vulnerability assessment methods. Such methods provide numerical values for the vulnerability of the operation of the entire molecular system to the dysfunction of each individual molecule. A high vulnerability for a molecule means that with high probability, the whole signaling network does not operate correctly, if that particular molecule is dysfunctional. Therefore, one can say such a molecule plays a key role in changing the state of the system from normal to faulty (abnormal). Identification of such important molecules, i.e., with high vulnerability levels, in signaling networks of interest is a major step towards understating the molecular basis of complex human diseases. From the drug development perspective, vulnerability assessment provides a set of candidate molecules to target.

To calculate the amount of the vulnerability of a network of interconnected pathways to the dysfunction of each molecule, one needs to consider a model for the network. There are different types of models such as differential equations, algebraic, Boolean, Bayesian, graphs, etc. [4]. We choose the simple yet powerful and biologically-appealing Boolean framework, where each molecule is either active (on) or inactive (off). Analogous to digital electronic circuits, one can say the state of a molecule is either 1 or 0, respectively. The Boolean framework and binary logic has been extensively used in the literature, to explore different characteristics of signaling and genetic networks [4][5]. In what follows, we demonstrate that inspired by digital circuit fault and reliability modeling engineering methods [2] and Boolean models of networks, one can compute the molecular system vulnerability to the dysfunction of each molecule within the system.

III. Vulnerability Analysis of Molecular Networks

Here we provide a summary of the proposed molecular vulnerability assessment algorithm [6], through which the vulnerability of a cellular signaling network to the dysfunction of its components can be calculated

1) Specify the inputs nodes (such as ligands, receptors, secondary messengers, etc.) and the output node (such as different transcription factors relevant to the input signal), as well as the intermediate molecules that allow the input signals to propagate from the inputs to the output. Then specify the type of the interactions among the molecules (stimulatory or inhibitory), using the existing literature.

2) Use Rule #1 and Rule #2 [6], to derive a binary logic equation for every intermediate molecule and the output molecule, using the interactions specified in Step 1. According to Rule #1, if a molecule has no inhibitory input, then the activation of at least one of its activatory inputs is enough to activate the molecule. On the other hand, based on Rule #2, if a molecule has at least one inhibitory input, then that molecule will be inactive, if at least one of its inhibitory inputs is activated (the molecule can be active, only if all the inhibitory inputs are inactive).

3) Construct the digital circuit of the network from the binary logic equations of Step 2, using the AND, OR, NOT and BUFFER digital circuit elements.

4) Identify the feedback paths of the digital circuit of Step 3, using the depth-first search (DFS) algorithm [7]. Then insert a flip-flop in each feedback path. If there is no feedback path, proceed directly to the next step.

5) Finally, apply the error probability propagation (EPP) algorithm [8] [9] to the circuit obtained in Step 4, to calculate the vulnerability levels of all the input and intermediate nodes (the vulnerability of the output node is always 1, since if the output node is dysfunctional the network will not operate efficiently anyway).

This multi-step vulnerability assessment algorithm is applied to several molecular networks such as a network leading to caspase3 activation and apoptosis (programmed cell death), and a network that controls the activation of p53, a tumor suppressor that is a transcriptional activator of several genes that ultimately control cell cycle arrest, cellular senescence, or apoptosis. The critical molecules identified by the proposed vulnerability assessment method are biologically relevant and are consistent with the experimental data [6]. In the next section we assess the vulnerability of the mitosis process to the dysfunction of molecules involved in the process.

IV. Vulnerability Analysis of the Mitosis Network

In this section we study the mitosis network, a signaling network in capillary endothelial cells that regulate cell cycle and is shown in Fig. 4a of [10]. The output node of the network represents the mitosis process, whereas input nodes are GF (growth factors) and spread, which represents the cell shape (spreading). The total number of interacting molecules and processes in the mitosis network is 11, with 18 interactions among them. The accuracy and usefulness of Boolean modeling approach to characterize the behavior of this network is experimentally confirmed in [10].

Based on the logic equations of this network [10], re-written in Table I using the notation of this paper, we have

TABLE I
Logic Equations of the Mitosis Network

Molecule	Logic equation
cyclinD1	cyclinD1=p27'×ERK
cyclinE	cyclinE=E2F×p27+E2F'×p27'
E2F	E2F=pRb'×cyclinE
ERK	ERK=(S phase genes)'×GF
Mitosis	Mitosis=Mitosis'×(S phase genes)
p27	p27=Mitosis×X+Mitosis'×X'
pRb	pRb=(cyclinD1×cyclinE)'
S phase genes	S phase genes=pRb'×E2F
X	X=GF×spread

Logic equations of the mitosis network. Each logic equation specifies the input signals to a molecule using the logic operations ', + and ×, which represent NOT, OR and AND, respectively. These equations are used to generate the digital electronic mitosis circuit in Fig. 1.

generated the mitosis digital circuit shown in Fig. 1, using AND, OR, BUFFER and NOT gates. The input (yellow)

nodes are GF (growth factors) and spread, which represents the cell shape (spreading), and the output (blue) node denotes the mitosis process. The node X stands for processes responsible for transducing the shape signal into a biochemical signal response.

There are five feedback paths in this circuit, identified using the DFS algorithm [7] and initiated from nodes cyclinD1, E2F, mitosis, and the NOTs of p27 and pRb. Following the well-known synchronous modeling approach in biological Boolean networks [4] [5], the nodes in the circuit are updated in a synchronous manner. This means that the 0/1 states of all the nodes change simultaneously, in discrete time steps. To implement this using the language of digital circuit engineering, one flip-flop is inserted in each feedback path. A flip-flop is a unit-delay digital logic circuit component, shown by FF in Fig. 1. By applying the EPP

TABLE II
VULNERABILITY VALUES OF THE MITOSIS CIRCUIT

Node	Vulnerability
Mitosis	1
p27	0.786
X	0.786
S phase genes	0.54
pRb	0.526
GF	0.445
cyclinE	0.396
E2F	0.388
spread	0.374
cyclinD1	0.183
ERK	0.155

These vulnerabilities are calculated using the EPP algorithm and are sorted from high to low. Note that p27 and X show the highest vulnerability levels.

method [8] [9] to the mitosis circuit in Fig. 1, derived from the logic equations of this network in Table I, we computed the vulnerability values of all the molecules in the mitosis network, listed in Table II.

The molecules in Table II. are sorted based on their vulnerability levels. Among different molecules in the mitosis pathway, p27 shows the highest vulnerability. This observation is also biologically and pathologically relevant since the critical role of p27 in regulation of cell cycle has been reported and it is now widely accepted that reduced p27 levels directly correlate with tumor aggressiveness and poor patient survival [11].

V. VULNERABILITY ANALYSIS OF THE CREB NETWORK

In this section a central nervous system network is considered that regulates the activity of the transcription factor CREB (cAMP responsive element binding). This transcription factor controls important functions of the cognitive and executive human brain, such as learning and memory.

We have constructed this network following the same approach as [12], to construct large molecular networks. The output node is the transcription factor CREB and the input nodes are seven major ligands in nervous system including glutamate, dopamine, GABA, serotonin, Ach, adenosine and enkephalin. The total number of interacting molecules in the CREB network in Fig. 2 is 64, with 152 inter-molecular interactions among them [6]. The logic equations for the CREB network are derived using Rule #1 and Rule #2 [6], and the corresponding digital electronic circuit, i.e., the CREB circuit, is shown in [6]. Using the proposed

TABLE III
VULNERABILITY VALUES OF THE CREB CIRCUIT

Node	Vul.	Node	Vul.	Node	Vul.
CREB	1	RSK	0.12	A2AR	0
Calmodulin	0.74	Adenosine	0.05	CaMKI	0
Ca^{2+}	0.6	Serotonin	0.05	CBP	0
cAMP	0.59	5-HT1AR	0.05	cJun	0
Galphai	0.58	A1R	0.045	DAG	0
AC2	0.58	Dopamine	0.04	5-HT2AR	0
AC1	0.57	Ach	0.035	5-HT4R	0
AC5	0.57	Enkephalin	0.035	GABABR	0
PKA	0.5	mGluR7	0.035	Galphas	0
P/Q type CaCh	0.5	D1R	0.025	Grb2	0
PP2A	0.5	D2R	0.025	GSK3	0
Gbetagamma	0.43	D3R	0.025	ILK	0
CaMKII	0.43	DOR	0.025	M1R	0
PP2B	0.27	KOR	0.025	PIP2	0
CaMKIV	0.27	M2R	0.025	AKT	0
CaMKK	0.27	M4R	0.025	PKC	0
CREM	0.27	MOR	0.025	PLCβ	0
N-type CaCh	0.25	NOR	0.025	PLCγ	0
NMDAR	0.24	Galphaz	0.02	RasGAP	0
PI3K	0.12	Glutamate	0.01	SAM68	0
PIP3	0.12	mGluR1	0.01		
PDK1	0.12	GABA	0		

vulnerability assessment algorithm, the vulnerabilities of all the molecules are calculated and listed in Table III.

The molecules that their dysfunction can result in the failure of the CREB function, with a probability of more than 0.5, are calmodulin, calcium, cAMP, Galphai, AC1, AC2, AC5, PKA, P/Q type calcium channel and PP2A. Elements of the cAMP-dependent signaling as well as some elements of calcium signaling molecules exhibit the highest vulnerability values in the CREB circuit. Furthermore, the distribution of the vulnerability values in the CREB network is highly non-uniform. On the other hand, the majority of the molecules, 41 out of 64, do not contribute to the failure of CREB circuit (vulnerability values less than 0.1).

The vulnerable molecules we have identified in this network are closely related molecules, and are also physiologically relevant to the CREB function. The name CREB was originally chosen because the discovered protein was a cAMP Responsive Element Binding protein. It is noteworthy that our engineering analysis has also identified cAMP and the molecules directly related to cAMP's function, such as AC1, AC2, AC5 and PKA, as the most critical molecules for the regulation of CREB. The crucial role of the cAMP-dependent kinase in the regulation of CREB has been experimentally established for many years now. A number of the most important aspects of the

cognitive and executive human brain, such as learning and memory, are known to be directly regulated by cAMP-dependent CREB functions [13]. In pathological terms, direct evidence for deregulation of PKA signaling has been reported in human disorders characterized by memory dysfunction such as Alzheimer Disease [14] or schizophrenia [15]. Vulnerability assessment of the CREB circuit has also identified several elements of calcium signaling to play a major role in CREB functioning. This observation is also physiologically and pathologically relevant, since the role of calcium signaling in regulation of CREB dependent functions is now well established [16], which is experimentally shown to have a crucial role in development of a number of pathological conditions raised from deregulation of calcium dependent signaling [17].

The proposed engineering method and its findings are verified via a variety of molecular biology experiments [6].

VI. Conclusion

In this paper a new systems biology framework is developed, inspired by the concepts of electronic circuit fault diagnosis engineering, to identify the vulnerable molecules that play crucial roles in the dysfunction of molecular networks. The vulnerable molecules we have identified are functionally related sets of molecules, with physiological and pathological relevance to the specific function of each network. This indicates the usefulness of the proposed methodology. This computer engineering approach is capable of identifying the critical molecules that have causative effects in human disorders and are promising targets for drug development [18].

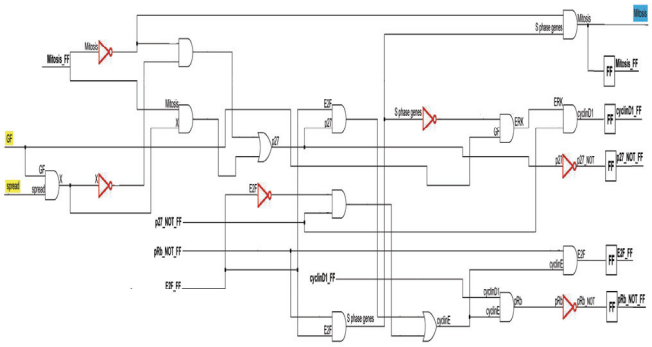

Fig. 1. The digital electronic mitosis circuit. The input (yellow) nodes are GF (growth factors) and spread, which represents the cell shape (spreading), and the output (blue) node denotes the mitosis process. The node X stands for processes responsible for transducing the shape signal into a biochemical signal response [10]. There are five feedback paths in this circuit, initiated from cyclinD1, E2F, mitosis, and the NOTs of p27 and pRb. These feedbacks are identified using the DFS algorithm, and then one flip-flop (FF) is inserted in each path. For example, there is an FF at the lower right corner of the circuit, with E2F and E2F_FF as its input and output, respectively. The name E2F_FF is appeared again at the lower left corner of the circuit, which means that E2F is "fed back" from the right side of the circuit to the left side. The feedback wire itself is not shown, to make the circuit diagram easier to read (figure can be zoomed in).

References

[1] T. Finkel and J. S. Gutkind, Eds. *Signal Transduction and Human Disease*. New York: Wiley, 2003.

[2] M. L. Bushnell and V. D. Agarwal, *Essentials of Electronic Testing for Digital, Memory and Mixed-Signal VLSI Circuits*. Boston, MA: Kluwer Academic Press, 2000.

[3] H. M. Sauro and B. N. Kholodenko, *Prog. Biophys. Mol. Biol.* vol. 86, p. 5, 2004.

[4] E. Klipp et al., *Systems Biology in Practice: Concepts, Implementation and Application*. Weinheim, Germany: Wiley-VCH, 2005.

[5] S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*. New York: Oxford University Press, 1993.

[6] A. Abdi, M. B. Tahoori and E. S. Emamian, "Fault diagnosis engineering of digital circuits can identify vulnerable molecules in complex cellular pathways," *Science Signaling*, vol. 1, p. 48, 2008.

[7] T. H. Cormen et al., *Introduction to Algorithms*, 2nd ed., MIT Press & McGraw-Hill, 2001.

[8] H. Asadi and M. B. Tahoori, in *Proc. IEEE International Symposium on Defect and Fault Tolerance of VLSI*, 2005, p. 463.

[9] H. Asadi and M. B. Tahoori, in *Proc. IEEE International Conference on Circuits and Systems*, 2005, p. 2991.

[10] S. Huang and D. E. Ingber, *Exp. Cell Res.*, vol. 261, p. 91, 2000.

[11] S. W. Blain and J. Massague, *Nat. Med.*, vol. 8, p. 1076, 2002.

[12] A. Ma'ayan et al., *Science*, vol. 309, p. 1078, 2005.

[13] B. E. Lonze and D. D. Ginty, *Neuron*, vol. 35, p. 605, 2002.

[14] G. A. Jicha et al., *J Neurosci.*, vol. 19, p. 7486, 1999.

[15] J. K. Millar et al., *Science*, vol. 310, p. 1187, 2005.

[16] A. E. West et al., *Proc. Natl. Acad. Sci. USA*, vol. 98, p. 11024, 2001.

[17] H. Tu et al., *Cell*, vol. 126, p. 981, 2006.

[18] J. Luo et al., *Cancer Cell*, vol. 4, p. 257, 2003.

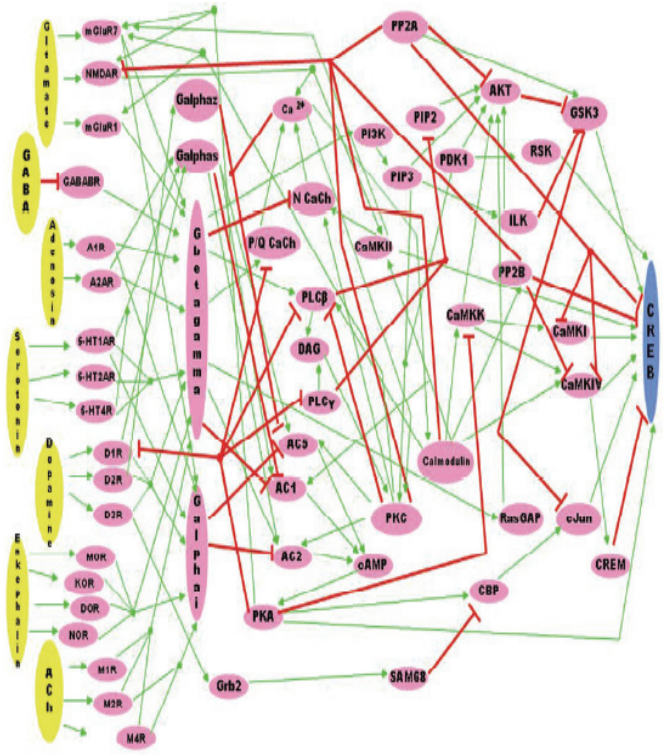

Fig. 2. The CREB network. It has a total of 64 molecules. The input (yellow) molecules are glutamate, dopamine, GABA, serotonin, Ach, adenosine and enkephalin, and the output (blue) molecule is CREB. To make the figure less crowded, small green and red circles are used to show that a signal coming from one molecule may target several other molecules. Also small green and red squares are employed to illustrate that several different molecules may target one molecule (figure can be zoomed in).