University of Wollongong

Research Online

1-1-2005

Gene network analysis and design

Matthew J. Berryman
University of Wollongong, mberryma@uow.edu.au

Andrew Allison
University of Adelaide

Derek Abbott
University of Adelaide, dabbott@eleceng.adelaide.edu.au

Recommended Citation

Gene network analysis and design

Matthew J. Berryman[a], Andrew Allison[a] and Derek Abbott[a]

[a]Centre for Biomedical Engineering and
School of Electrical and Electronic Engineering,
The University of Adelaide, SA 5005, Australia.

ABSTRACT

Gene networks are composed of many different interacting genes and gene products (RNAs and proteins). They can be thought of as switching regions in n-dimensional space or as mass-balanced signaling networks. Both approaches allow for describing gene networks with the limited quantitative or even qualitative data available. We show how these approaches can be used in modeling the apoptosis gene network that has a vital role in tumor development. The open question is whether engineering changes to this network could be used as a possible cancer treatment.

Keywords: gene networks, signaling, apoptosis, cancer

1. INTRODUCTION

While cancer is understood in very broad terms[1,2] and many of the key proteins involved are very well understood,[3–5] more research is needed at the gene network level to understand interactions between genes, proteins, and the environment. p53 is a key protein involved in a number of key cell processes,[6] which if disturbed can lead to cancer.[1,2] By modeling the p53 gene network, we can gain a better understanding of its interactions, with key implications for cancer treatment.

Modeling gene networks gives us a broad, yet important view of vital cell regulatory functions.[7,8] Two techniques that have been developed are mass-balanced signaling networks[9–11] (of which the Michaelis-Menten kinetics are a particular form[12]), and piecewise-linear differential equations (DEs),[7,13,14] a generalization on the work in Boolean NK switching networks by Glass and Edwards.[14–16] Although enzyme kinematics are a useful tool,[9,12] the ODEs involved are typically quite stiff which increases the computation time.[10] Also, the piecewise-linear DE approach gives a better picture of the qualitative nature of the behaviour,[7] since they are quick and accurate to integrate[14] and good analytical tools exist for their qualitative behaviour.[13]

Gene networks present several problems for any sort of quantitative analysis.[7,13] Many of the interacting proteins may be unknown, and even between the known interacting proteins, some of the interactions occur only under unknown conditions.[7] Furthermore, although techniques such as microarrays and reverse transcriptase-polymerase chain reaction (RT-PCR) give gene expression levels,[17] there is a lot of variance, a lot of measurement error, and they do not typically provide enough detail about the complex interactions to establish any strict rules of behaviour in the gene networks.[18] Even in well-established gene networks, such as the p53 network, the details of many interactions are still being discovered.[6] Some of the key ones, such as the feedback loop that exists between p53 and another protein, Mdm2, are lacking clear quantitative values.[6] Consequently, we are left with an incomplete picture of protein interactions, with imprecise, qualitative rules – for example, protein A binds to protein B, or protein C activates transcription of protein D. Basing our analysis on logical switching systems is useful, because:

"*logical switching systems* capture major features of a homologous class of nonlinear dynamical systems governed by sigmoidal functions because such systems tend to *sharpen* their responses to *extremal values of the variables*." (Kauffman[8])

Send correspondence to Derek Abbott
E-mail: dabbott@eleceng.adelaide.edu.au, Telephone: +61 8 8303 5748

If one considers a typical sigmoidal function,

$$y = \frac{x^n}{\theta^n + x^n},\tag{1}$$

then it is trivial to show that

$$\lim_{n\to\infty} y = \lim_{n\to\infty} \frac{\frac{x^n}{\theta_n}}{1 + \frac{x^n}{\theta^n}} = \begin{cases} 1, & x > \theta \\ 0, & x < \theta. \end{cases}\tag{2}$$

Thus in the extreme case, as $n \to \infty$ we have a simple switching, or step function. Piecewise-linear (PL) differential equation (DE) models, originally developed by Glass and Edwards,[14–16] offer a way around this problem, in that they can successfully model systems in a qualitative sense, including quantitative data where available. Filippov successfully introduced step-functions, which are important for modeling the switch-like regulatory interactions in gene networks.[19]

2. METHODS

As described by de Jong,[13] the rates of change of proteins can be written as

$$\dot{\mathbf{x}} = \mathbf{f}\left(\mathbf{x}\right) - \mathbf{g}\left(\mathbf{x}\right)\mathbf{x}\tag{3}$$

where $\mathbf{x} = (x_1, \ldots, x_n)'$, $\mathbf{x} \in \mathbb{R}_{\geq 0}^n$ is the expression levels (copy number / volume) of the n proteins, $\mathbf{f} = (f_1, \ldots, f_n)'$ consists of the rates of synthesis of each proteins, and $\mathbf{g}\left(\mathbf{x}\right) = \mathrm{diag}\left(g_1, \ldots, g_2\right)$ is the rate of degradation.[13] The rate of synthesis of the protein i, $f_i : \mathbb{R}_{\geq 0} \to \{0, 1\}$ depends on the levels of the proteins, $\mathbf{x}$ in some fashion,

$$f_i\left(\mathbf{x}\right) = \sum_{l \in \mathcal{L}} \kappa_{il} b_{il}\left(\mathbf{x}\right),\tag{4}$$

for rate parameter $\kappa_{il} > 0$, and $b_{il} : \mathbb{R}_{\geq 0}^n \to \{0, 1\}$ a regulation function, and $\mathcal{L}$ a possibly empty set of indices of *regulation functions*. The g_i are defined similarly to the f_i, except we require that $g_i > 0 \forall i$. The *regulation functions* b_{il} are defined in terms of step functions $\mathrm{s}^+, \mathrm{s}^- : \mathbb{R}^2 \to \{0, 1\}$ defined as

$$\mathrm{s}^+\left(x_j, \theta_j\right) = \begin{cases} 1, & x_j > \theta_j \\ 0, & x_j < \theta_j \end{cases}\tag{5}$$

$$\mathrm{s}^-\left(x_j, \theta_j\right) = 1 - \mathrm{s}^+\left(x_j, \theta_j\right)\tag{6}$$

We consider a generalized view of the p53 network, as shown in Figure 1. Based on Figure 1, we abstract over the proteins apart from p53 and Mdm2, instead considering $\mathbf{x} = (x_d, x_u, x_p, x_m, x_A, x_C, x_D)'$ where x_d is the level of DNA damage kinases, x_u the level of UV stress kinases, x_p the expression level of p53, x_m the level of Mdm2, x_D the combined level of apoptosis (programmed cell death) proteins, x_A for the angiogenesis proteins, and x_C for the cell cycle proteins. The network of interactions is shown in Figure 2. We use the following set of equations to model the p53 network:

$$x_p = s_{pd}\,\mathrm{s}^+\left(x_d, \theta_d\right) + s_{pu}\,\mathrm{s}^+\left(x_u, \theta_u\right) + s_{pm}\,\mathrm{s}^+\left(x_m, \theta_m\right)\mathrm{s}^-\left(x_p, \theta_p\right) - g_p x_p\tag{7}$$

$$x_m = s_m\,\mathrm{s}^+\left(x_m, \theta_m\right)\mathrm{s}^-\left(x_p, \theta_p\right) - g_m x_m\tag{8}$$

$$x_A = s_A\,\mathrm{s}^+\left(x_p, \theta_A\right) - g_A x_A\tag{9}$$

$$x_C = s_C\,\mathrm{s}^+\left(x_p, \theta_C\right) - g_C x_C\tag{10}$$

$$x_D = s_D\,\mathrm{s}^+\left(x_p, \theta_D\right) - g_D x_D,\tag{11}$$

where θ_a for some protein (or protein group) a is the level at which the protein b in the term $s_{cb}\,\mathrm{s}^{\pm}\left(x_c, \theta_a\right)$ signals for the synthesis of the protein c at rate s_{cb}. No equations for x_d and x_u are given as these are input variables. We define $\theta_C < \theta_A < \theta_D$, which has the interpretation that as a first response the cell stops the cell cycle and
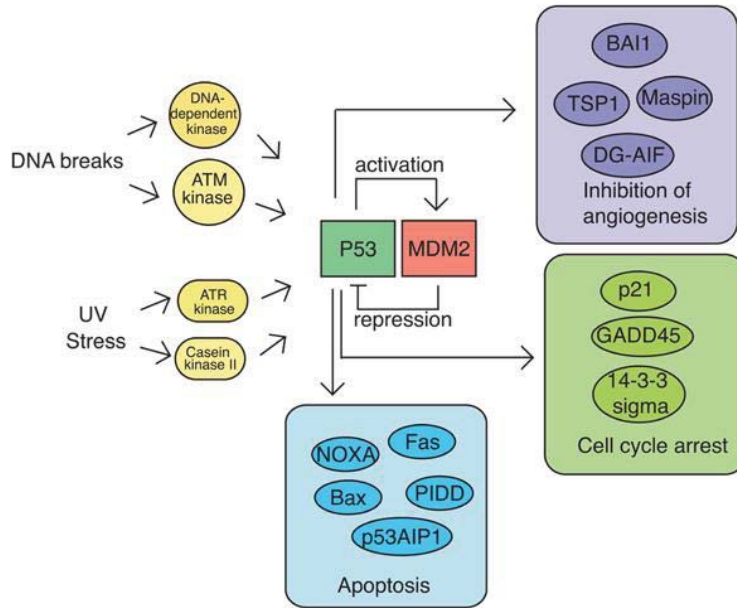
Figure 1. This figure, by Nakade,[20] shows some of the key players in the p53 network, including the p53-Mdm2 feedback loop. DNA breaks and UV stress cause changes in such kinase proteins as ATM kinase and Casein kinase II. A kinase is a protein that adds a phosphor atom to a protein. In this case, a phosphor atom is added to p53. p53 is then termed *phosphorylated*, and this affects its binding to Mdm2, altering the feedback loop. The increased levels of p53 then can signal for inhibition of angiogenesis, cell cycle arrest (to allow DNA repair to occur), and/or apoptosis. The actual response, simply halting or undergoing apoptosis, is a function of the input.

signals for angiogenisis, and with the highest levels of damage and stress undergoes apoptosis (programmed cell death).

For each output variable, we have a target equilibrium position, which exists in some regulatory domain. Switching domains exist on the boundaries between regulatory domains. This can be seen in the representation of the domains for p53 and Mdm2 in Figure 3.

The equilibrium positions and the domains are shown in Table 1. Note that m_a is the maximum value of expression of protein a, denoted by x_a. For the output protein levels of interest, we also define critical values – if the protein level x_A lies above the critical value c_A, then we say protein group A is activated, that is the cell would then be signaling for angiogenesis. These are the primary domains used for our analysis, however we have an alternate set to see what happens if we can reduce the level of p53 produced in response to various stress levels. These are shown in Table 2. We also need a set of initial conditions. Because this is a qualitative model, it is sufficient to simply specify regulatory domains for the initial conditions. We are interested in four different cases,

1. $(x_d, x_u) \in [0, \theta_d) \times [0, \theta_u)$,

2. $(x_d, x_u) \in [0, \theta_d) \times (\theta_u, m_u]$,

3. $(x_d, x_u) \in (\theta_d, m_d] \times [0, \theta_u]$, and

4. $(x_d, x_u) \in (\theta_d, m_d] \times (\theta_u, m_u]$,

corresponding to varying stresses on the cell; the first one corresponding to no DNA damage or UV stress. In addition, we must also specify the initial conditions for the other variables, these are kept constant as shown in Table 3.

$x \in (a, b)$ means $a < x < b$, $x \in (a, b]$ means $a < x \leq b$, $x \in [a, b)$ means $a \leq x < b$ and $x \in [a, b]$ means $a \leq x \leq b$
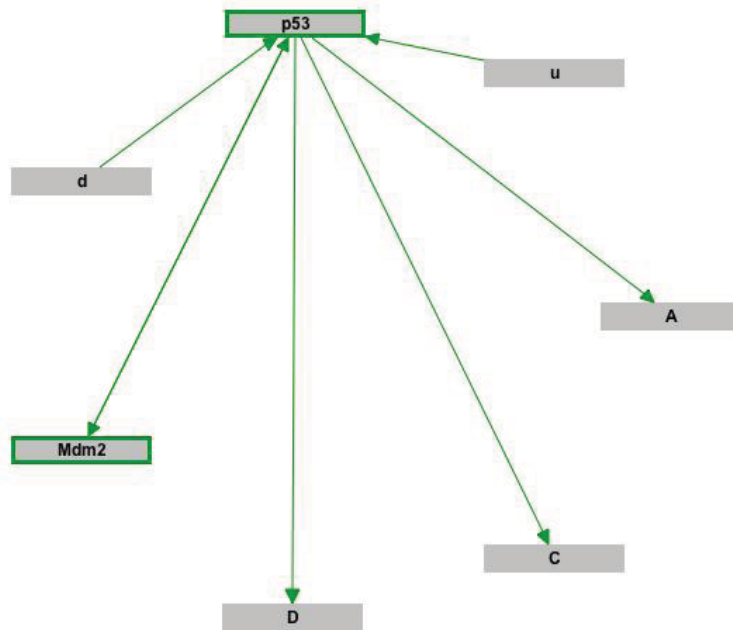
Figure 2. This network shows with directed arrows the influences of various inputs and proteins on other proteins. UV stress (denoted by u) and DNA damage (d) signal for production of p53, which is in a feedback loop with the Mdm2 protein. Downstream targets of p53 include proteins that signal for angiogeneis (a), halting of the cell cycle (c), and apoptosis (programmed cell death, denoted by d).
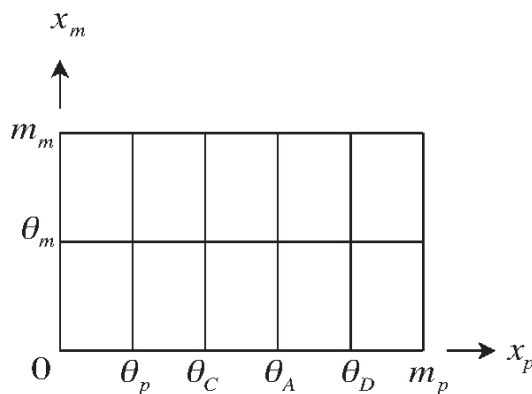
Figure 3. The set of protein expression values $(x_p, x_m) \in [0, m_p] \times [0, m_m]$ can be drawn as a 2D plane, with points lying either in switching domains, the lines at critical values (θ's), and the regulatory domains (lying between critical values). As the level of p53 (x_p) increases, it shifts in to different signaling domains. For example if x_p lies in the range (θ_C, θ_A) it means that cell-cycle halting is being signaled for, but not angiogenesis or apoptosis.

Table 1. The equilibrium position(s) for each output protein level, and the regulatory domains that have been specified for these to occur in. There are several p53 equilibrium positions – one for each combination of synthesis/binding rates, the others have only one synthesis rate so only one equilibrium position occurs, according to the theorems in de Jong.[13]

Variable	Equilibrium position	Regulatory Domain
x_p	$(s_{pd} + s_{pu})/g_p$	$(\theta_D, m_m]$
x_p	$(s_{pd} + s_{pu} + s_{pm})/g_p$	$(\theta_D, m_m]$
x_p	s_{pd}/g_p	(θ_A, θ_D)
x_p	$(s_{pd} + s_{pm})/g_p$	$(\theta_D, m_m]$
x_p	s_{pu}/g_p	(θ_A, θ_D)
x_p	$(s_{pu} + s_{pm})/g_p$	$(\theta_D, m_m]$
x_p	s_{pm}/g_p	(θ_A, θ_D)
x_m	s_m/g_m	$(\theta_m, m_m]$
x_C	s_C/g_C	$(c_C, m_C]$
x_A	s_A/g_A	$(c_A, m_A]$
x_D	s_D/g_D	$(c_D, m_D]$

Table 2. Similar to Table 1, we list the lowered equilibrium position(s) for each output protein level, and the regulatory domains that have been specified for these to occur in. As discussed previously, p53 has several equilibrium positions – one for each combination of synthesis/binding rates, the others have only one synthesis rate so only one equilibrium position occurs, according to the theorems in de Jong.[13]

Variable	Equilibrium position	Regulatory Domain
x_p	$(s_{pd} + s_{pu})/g_p$	(θ_A, θ_D)
x_p	$(s_{pd} + s_{pu} + s_{pm})/g_p$	$(\theta_D, m_m]$
x_p	s_{pd}/g_p	(θ_C, θ_A)
x_p	$(s_{pd} + s_{pm})/g_p$	(θ_A, θ_D)
x_p	s_{pu}/g_p	(θ_C, θ_A)
x_p	$(s_{pu} + s_{pm})/g_p$	(θ_A, θ_D)
x_p	s_{pm}/g_p	(θ_C, θ_A)
x_m	s_m/g_m	$(\theta_m, m_m]$
x_C	s_C/g_C	$(c_C, m_C]$
x_A	s_A/g_A	$(c_A, m_A]$
x_D	s_D/g_D	$(c_D, m_D]$

Table 3. Initial regulatory domains for p53 (x_p), Mdm2 (x_m), and output signals for the halting of cell cycle (x_C), angiogenesis (x_A), and apoptosis (x_D).

Variable	Initial domain
x_p	$[0, \theta_A]$
x_m	$[0, m_m]$
x_C	$[0, c_C]$
x_A	$[0, c_A]$
x_D	$[0, c_D]$

Figure 4. Graph showing the final set of states for the input set one, $(x_d, x_u) \in [0, \theta_d) \times [0, \theta_u)$, representing no external stress on the cell. States denote the proteins being in particular domains, for example Tables 4 denotes the final state S1. The numbering of states is arbitrary. From the initial equilibrium state, the cell can transition through a number of regulatory (boxed) and switching (not boxed) states and lying in an attractor basis with the final equilibrium state in Table 4 as the attractor. Arrow directions denote allowed state transitions.

3. RESULTS

Each set of initial conditions run through the model produces a set of states that the cell cycles through. We found that in each case there is a single attractor state that the system settles into, representing the decision of the cell to either continue normally, halt the cell cycle, signal for angiogenesis, or undergo apoptosis. An example of the transition diagram is shown in Figure 4. The set of four states, corresponding to the final states for each input 1-4, and the normal equilibrium postions in Table 1 are shown in Tables 4 through 7.

Table 4. Final state for input set one: $(x_d, x_u) \in [0, \theta_d) \times [0, \theta_u)$. This represents no external stress on the cell. As expected, none of its responses are triggered, represented by $x_a < c_A$, $x_C < c_C$ and $x_D < c_D$.

Variable	Final domain
x_A	$[0, c_A)$
x_C	$[0, c_C)$
x_D	$[0, c_D)$
x_m	$[0, \theta_m)$
x_p	$[0, \theta_p)$

For the alternate set of equilibrium positions in Table 2, the results are the same, except that when a single input is present, the system only signals for the cell cycle to be halted and not angiogensis, and making $c_A < c_C$

Table 5. Final state for input set two: $(x_d, x_u) \in [0, \theta_d] \times (\theta_u, m_u]$. This represents one external stress on the cell, in this case UV stress. Two of its responses are triggered, represented by $c_A < x_a \leq m_A$, $c_C < x_C \leq m_C$ and $c_D < x_D \leq m_D$. In words, this means the cell is signaling for angiogenesis, and has halted the cell cycle in order to prevent DNA damage, but the cell is not under enough pressure to undergo apoptosis.

Variable	Final domain
x_A	$(c_A, m_A]$
x_C	$(c_C, m_C]$
x_D	$[0, c_D)$
x_m	θ_m
x_p	(θ_A, θ_D)

Table 6. Final state for input set three: $(x_d, x_u) \in (\theta_d, m_d] \times [0, \theta_u]$. This represents one external stress on the cell, in this case DNA damage. Two of its responses are triggered, represented by $c_A < x_a \leq m_A$, $c_C < x_C \leq m_C$ and $c_D < x_D \leq m_D$. In words, this means the cell is signaling for angiogenesis, and has halted the cell cycle to repair DNA damage, but the cell is not under enough pressure to undergo apoptosis.

Variable	Final domain
x_A	$(c_A, m_A]$
x_C	$(c_C, m_C]$
x_D	$[0, c_D)$
x_m	θ_m
x_p	(θ_A, θ_D)

Table 7. Final state for input set four: $(x_d, x_u) \in (\theta_d, m_d] \times (\theta_u, m_u]$. This represents one external stress on the cell, in this case DNA damage. Two of its responses are triggered, represented by $c_A < x_a \leq m_A$, $c_C < x_C \leq m_C$ and $c_D < x_D \leq m_D$. In words, this means the cell is so stressed and damaged that it would be too risky for cell division to occur, which would fix in the genetic changes to its descendants, so it undergoes apoptosis.

Variable	Final domain
x_A	$(c_A, m_A]$
x_C	$(c_C, m_C]$
x_D	$[c_D, m_D)$
x_m	θ_m
x_p	$(\theta_A \theta_D)$

reverses that. Also for both inputs being present, the output is that both the cell cycle and angiogenesis are signaled for, but not apoptosis. This allows potentially harmful mutations to be passed on to daughter cells.

4. CONCLUSIONS

Given a qualitative model of the interactions of the p53 protein in the cell, we are able to design a simplified gene network to represent this information. In a qualitative sense this gives the output that is observed in *in vivo* and *in vitro* cells, namely that if not stressed or damaged, the cell undergoes a normal cell cycle and divides. If there is some damage or stress the cell cycle is halted, and for a significant amount of damage and stress the cell does not divide and pass on its mutations but instead undergoes apoptosis (programmed cell death). Shifting the equilibrium positions is equivalent to altering the response of the cell; for the example we considered, changing the equilibrium positions lower means the cell is less likely to signal for appropriate responses, in particular never undergoing apoptosis. This would mean that potentially harmful (to the organism) mutations are carried forward into daughter cells upon division of the stressed and damaged cell. Shifting the equilibrium positions higher means the cell is more likely to halt and repair DNA and undergo apoptosis for higher (yet lower than normal) levels of DNA damage. This could be a useful response to have in tumors that have not lost the capability of apoptosis and suggests future treatments could target these equilibrium positions.

In future work, we propose exploring the angiogenesis, cell cycle, and apoptosis pathways in more detail.

The use of quantitative data where available would allow us to build testable models of cell outputs, although our existing model agrees with observed qualitative descriptions of cell processes. In addition to exploring more equilibrium positions and inputs, we can also explore changes to the network. Since many drugs alter the pathways, we could possibly impact upon drug selection, dosage, and development of new drugs for treating cancer.

Acknowledgments

REFERENCES

1. S. L. Spencer, M. J. Berryman, J. García, and D. Abbott, "An ordinary differential equation model for the multistep transformation to cancer," *Journal of Theoretical Biology* **231**, pp. 515–524, 2004.
2. D. Hanahan and R. Weinberg, "The hallmarks of cancer," *Cell* **100**, pp. 57–70, 2000.
3. P. A. Futreal, L. Coin, M. Marshall, T. Down, T. Hubbard, R. Wooster, N. Rahman, and M. R. Stratton, "A census of human cancer genes," *Nature Reviews Cancer* **4**, pp. 177–183, 2004.
4. J. W. Arends, "Molecular interactions in the Vogelstein model of colorectal carcinoma," *Journal of Pathology* **190**, pp. 412–416, 2000.
5. A. G. Knudson, "Cancer genetics," *American Journal of Medical Genetics* **111**, pp. 96–102, 2002.
6. B. Vogelstein, D. Lane, and A. J. Levine, "Surfing the p53 network," *Nature* **408**, pp. 307–310, 2000.
7. H. de Jong, J. Geiselmann, G. Batt, C. Hernandez, and M. Page, "Qualitative simulation of the initiation of sporulation in *bacillus subtilis*," *Bulletin of Mathematical Biology* **66**, pp. 261–299, 2004.
8. S. A. Kauffman, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, 1993.
9. B. Aleman-Meza, Y. Yu, H.-B. Schüttler, J. Arnold, and T. R. Taha, "KINSOLVER: A simulator for compting large ensembles of biochemical and gene regulatory networks," in *http://lsdis.cs.uga.edu/ aleman/kinsolver/*, 2004.
10. H. Kurata, T. Inoue, Y. Sumida, S. Tanaka, and T. Ohashi, "Simulation and system analysis of gene regulatory networks using the two-phase partition method," *Genomes Informatics* **12**, pp. 286–287, 2001.
11. M. Kærn, W. J. Blake, and J. J. Collins, "The engineering of gene regulatory networks," *Ann. Rev. Biomed. Eng.* **5**, pp. 179–206, 2003.
12. J. D. Murray, *Mathematical Biology*, Springer, 3rd ed., 2002.
13. H. de Jong, J.-L. Gouzé, C. Hernandez, M. Page, T. Sari, and J. Geiselmann, "Quantitative simulation of genetic regulatory networks using piecewise-linear models," *Bulletin of Mathematical Biology* **66**, pp. 301–340, 2004.
14. R. Edwards and L. Glass, "Combinatorial explosion in model gene networks," *Chaos* **10**, pp. 691–704, 2000.
15. L. Glass, "Classification of biological networks by their qualitative dynamics," *Journal of Theoretical Biology* **54**, pp. 85–107, 1975.
16. R. Edwards, "Analysis of continuous-time switching networks," *Physica D* **146**, pp. 165–199, 2000.
17. D. D. Shoemaker, E. E. Schadt, C. D. Armour, Y. D. He, P. Garrett-Engele, P. D. McDonagh, L. P. M., *et al.*, "Experimental annotation of the human genome using microarray technology," *Nature* **409**, pp. 922–927, Feb. 2001.
18. Y. Moreau, F. de Smet, G. Thijs, K. Marchal, and B. de Moor, "Functional bioinformatics of microarray data: from expression to regulation," *Proceedings of the IEEE* **90**, pp. 1722–1743, Nov. 2002.
19. A. F. Filippov, *Differential equations with discontinuous righthand sides*, Kluwer Academic Publishers, 1988.
20. K. Nakade, "The p53 network," in *http://www.brc.riken.jp/lab/dna/en/GENESETBANK/p53_network.html*, 2004.