Creating Metabolic and Regulatory Network Models using Fuzzy Cognitive Maps

J.A. Dickerson and Z. Cox
Electrical Engineering Dept.,
Iowa State University
Ames, IA, USA
julied@iastate.edu

E. S. Wurtele
Botany Dept.
Iowa State University
Ames, IA, USA
mash@iastate.edu

A.W. Fulmer
Procter & Gamble Co.
Corporate Research
Cincinnati, OH, USA
fulmer.aw@pg.com

Abstract

This paper describes a model of metabolic networks that uses fuzzy cognitive maps. Nodes of the map represent specific biochemicals such as proteins, RNA, and small molecules, or stimuli, such as light, heat, or nutrients. Edges of the map capture regulatory and metabolic relationships found in biological systems. These relationships are established by a domain expert, the biological literature, and extracted from RNA microarray data. This work is part of the development of a software tool, *FCModeler*, which models and visualizes metabolic networks. A model of the metabolism of the plant hormone gibberellin in *Arabidopsis* is used to show the capabilities of the fuzzy model.

1. Introduction

Metabolic networks form the basis for the net accumulation of biomolecules in organisms. Regulatory networks modulate the action of metabolic networks, leading to physiological and morphological changes. A variety of data types will be integrated to learn about these pathways. These include global profiling of protein levels, RNA levels (microarrays), metabolite levels, and metabolic flux measurements. Such emerging technologies are yielding vast amounts of data on gene expression. This points to the need to develop more methodologies to identify and analyze complex biological networks.

Several types of models have been proposed for representing gene regulatory networks, including Boolean networks [1, 2], linear weighting networks [3], differential equations [4], and Petri nets [5]. Circuit simulations and differential equations require detailed information that is not yet known about the regulatory mechanisms between genes. Boolean networks analyze binary state transition matrices to look for patterns in gene expression. Each part of the network is either on or off depending on whether a signal exceeds a pre-determined threshold. Linear weighting networks have the advantage of simplicity since they use simple weight matrices to additively combine the contributions of different regulatory elements. Petri nets can handle a wide variety of information, however their complexity does not scale up well to systems that have both continuous and discrete inputs [6, 7]. All of these models are based on gene expression level information. However, gene expression levels alone cannot give a complete picture of how the metabolism of living things works [8].

Fuzzy cognitive maps (FCMs) have the potential to answer many of the concerns that arise from the existing models[9-11]. FCMs have been successfully applied to systems that have uncertain and incomplete models that cannot be expressed compactly or conveniently in equations. Examples include modeling human psychology [12] and on-line fault diagnosis at power plants [13]. These problems have several common features. The first is the lack of quantitative information on how different variables interact. The second is that the direction of causality is at least partly known and can be articulated by a domain expert. The third is that they link concepts from different domains together using arrows of causality. These features are shared by the problem of modeling metabolic networks. The FCModeler tool uses fuzzy methods for modeling networks and interprets the results using fuzzy cognitive maps. The FCModeler tool is intended to capture the intuitions of biologists, help test hypotheses, and provide a modeling framework for assessing the large amounts of data captured by RNA microarrays and other high-throughput experiments [14].

2. Structure of Concepts and Links

The nodes in the FCM represent specific biochemicals such as proteins, RNA, and small molecules, or stimuli, such as light, heat, or nutrients. There are three basic types of links specified. In a conversion link (black arrow, shown as a dotted line), a node (typically a chemical(s)) is converted into another node, and used up in the process. In a regulatory link (green and red

arrows, shown as solid arrows with a plus or minus sign), the node activates or deactivates another node, and is not used up in the process. A catalytic link (blue arrows, shown as a thick line) represents an enzyme that enables a chemical conversion and does not get used up in the process. Figure 1 shows a small part of a graph for the Arabadopsis metabolic and regulatory network.

In the metabolic network database, the type of link is further delineated by the link mechanism and the certainty. Some of the current mechanisms are: direct, indirect, and ligand. Direct links assume a direct physical interaction. Indirect links assume that the upstream node activates the downstream node indirectly and allows for the existence of intermediate nodes in such a path. The ligand link is a "second messenger" mechanism in which a node produces or helps produce a ligand (or small molecule that binds)
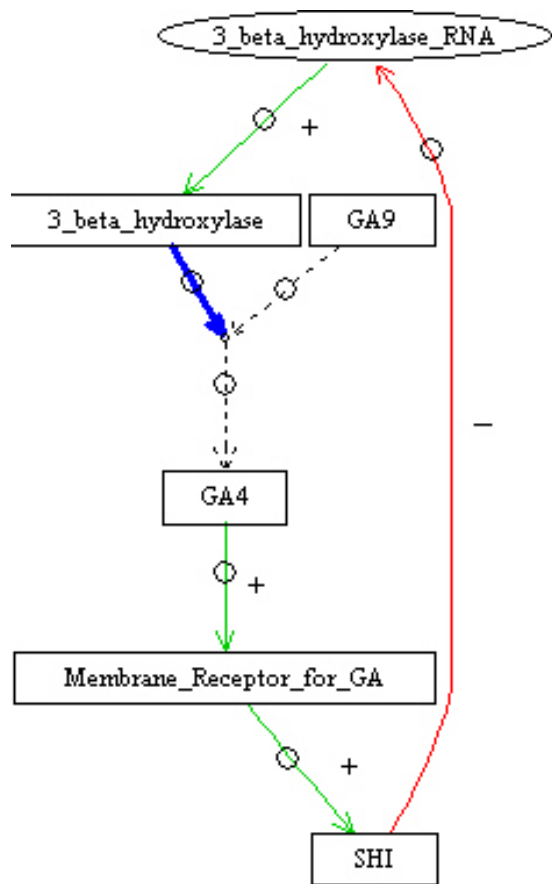

Figure 1. This is a map of a networking model of gibberellin (active form is GA4). The sequence is started by translationof 3_beta_hydroxylase_RNA into the 3_beta_hydroxylase protein. Dashed lines are conversion links, bold lines are catalytic links, and solid lines are regulatory links.

and either "activates" or "inhibits" a target node. Often the nature of the link is unknown and it cannot be modeled in the current framework. The link certainty expresses a degree of confidence about the link. This will be used for hypothesis testing.

3. Visualizing Metabolic Networks

The known and unknown biological information in the metabolic network is visualized using a graph visualization tool. The graph layout program is *dot* which is part of the *Graphviz* program developed at AT&T research labs (see http://www.research.att.com sw/tools/graphviz/. *Diva*, a software infrastructure for visualizing and interacting with dynamic information spaces (http://www.gigascale.org/diva/), is used for displaying the graphs. This allows for a greater variety of visualization objects on the display. The front end of the FCModeler tool is a Java [TM] interface that reads and displays data from a database of links and nodes.

4. Fuzzy Cognitive Map Modeling for Metabolic and Regulatory Networks

Fuzzy cognitive maps are fuzzy digraphs that model causal flow between concepts or, in this case, genes, proteins, and transcription factors [10, 11]. The concepts are linked by edges that show the degree to which the concepts depend on each other. FCMs can be binary state systems called simple FCMs with causality directions that are +1, a positive causal connection, -1, a negative connection, or zero, no causal connection. The fuzzy structure allows the gene or protein levels to be expressed in the continuous range [0, 1]. The input is the sum of the product of the fuzzy edge values. The system nonlinearly transforms the weighted input to each node using a threshold function or other nonlinear activation. FCMs are signed digraphs with feedback. Nodes stand for causal fuzzy sets where events occur to some degree. Edges stand for causal flow. The sign of an edge (+ or -) shows causal increase or decrease between nodes.

Each causal node $C_i(t)$ is a nonlinear function that maps the output activation into a fuzzy membership degree in [0,1]. Simple or trivalent FCMs have causal edge weights in the set {-1,0,1} and concept values in {0,1} or {-1,1}. Simple FCMs give a quick approximation to an expert's causal knowledge. More detailed graphs can replace this link with a time-dependent and/or nonlinear function.

FCMs recall as the FCM dynamical system equilibrates. Simple FCM inference is matrix-vector multiplication followed by thresholding. State vectors C_n cycle through the FCM edge matrix E, that defines the edges e_{ki} where k is the upstream node and i is the

downstream node. The system nonlinearly transforms the weighted input to each node C_i:

$$C_i(t_{n+1}) = S\left[\sum e_{ki}(t_n)C_k(t_n)\right] \tag{1}$$

$S(y)$ is a monotonic signal function bounded function such as the sigmoid function:

$$S_j(y_j) = \frac{1}{1+e^{-c(y_j-T_j)}} \tag{2}$$

In this case $c=1000$ and $T_j = 0.5$ for all nodes. This is equivalent to a step function with a threshold at 0.5. The edges between nodes can also be time dependent functions that create a complex dynamical system.

Regulatory Links: The regulatory edges are modeled using a simple FCM model that assumes binary connecting edges: $e_{ki} = \{-1,1\}$ for the single edge case. When there are multiple excitatory or inhibitory connections, the weights are divided by the number of input connections in the absence of other information. As more information becomes known about details of the regulation, for example how an RNA level affects the translation of the corresponding protein, the function of the link models will be updated. The regulatory nodes will also have self-feedback since the nodes stay on until they have been inhibited.

Conversion Links: Conversion relationships are modeled in different ways depending on the goal of the simulation study. The first case corresponds to investigating causal relationships between nodes. The node is modeled in the same manner as a regulatory link in which the presence of one node causes presence at the next node. When information about the rate of change in a reaction is available, a simple difference equation can model the gradually rising and falling levels of the nodes. The step size depends on the reaction rate and the stoichiometric relationship between the nodes.
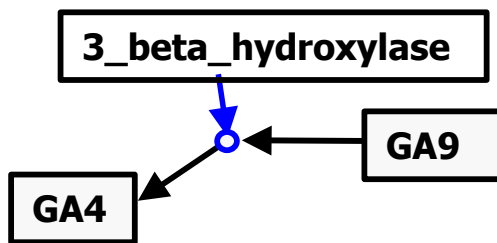

Figure 2. Catalyzing link. In this case, 3_beta_hydroxylase is the catalyst for a reaction between GA9 and GA4.

Catalyzed Links: Catalyzed reactions add a dummy node that acts upon a conversion link. This allows one link to modify another link. In this paper, the catalyzed link is simulated by weighting the inputs into the dummy node in such a way that both inputs much be present for the node to be active. Another method of modeling catalyzed links is an augmented matrix that operates on the edges between the nodes. Figure 2 gives an example of a catalyzing node model. The blue circle is a dummy node so the catalyst, 3 beta_hydroxylase can modify the reaction between GA4 and GA9.

Forcing functions: In biological systems such as cells, many of the metabolic network elements are always present. This is modeled as:

$$C_i(t_{n+1}) = S\left[\sum e_{ki}(t_n)C_k(t_n)+1\right]. \tag{3}$$

The effect of this equation is that if the node is not inhibited, it is active.

5. Results

The metabolism and signal transduction of the plant hormone gibberellin in Arabidopsis [15,16] was used to test this modeling scheme. Figure 1 shows a potential network configuration. The link types and causality directions were created by an expert researcher in the field. The key element in this graph is the block labeled GA4. This compound regulates many other regulatory mechanisms in plants. IPP, GAI, GRS, SPY, GA_MYB, FACTORX, LEC, and Meristem Carpal Margin had forcing functions applied to them. The results can be visualized using a time plot that shows which node was active and to what degree as shown in Figure 3.

Figures 4 through 7 show visualized networks at different time steps to analyze the interactions in the network. Figures 4-6 show the operation of the catalyzing node. Figure 7 shows how GA4 can regulate its own production and how it can control the activity of other biomolecules. Thus results a homeostatic control of GA4 levels. The oscillation of the GA levels directs the generation of biomolecules that, in the absence of other constraining factors, are implicated in the formation of new cellular proliferation centers, referred to as meristems. Many key features of this model, including timing, can be tested experimentally and relatively rapidly by globally monitoring temporal profiles of mRNA, protein and metabolite.

6. Conclusions

Metabolic and regulatory networks can be modeled using fuzzy cognitive maps. Future plans include: Simulating intervention in the network (e.g. what happens when a node is shut off), searching for critical paths and control points in the network, and capturing information about how edges between graph nodes change when different regulatory factors are present.

Acknowledgement

References:

[1] S. Liang, S. Fuhrman, and R. Somogyi, "REVEAL, A general reverse engineering algorithm for inference of genetic network architectures," presented at Pacific Symposium on Biocomputing 3, Hawaii, 1998.

[2] T. Akutsu, S. Miyano, and S. Kuhara, "Identification of Genetic Networks from a Small Number of Gene Expression Patterns Under the Boolean Network Model," presented at Pacific Symposium on Biocomputing 4, Hawaii, 1999.

[3] D. C. Weaver, C. T. Workman, and G. D. Stormo, "Modeling Regulatory Networks with Weight Matrices," presented at Pacific Symposium on Biocomputing 4, Hawaii, 1999.

[4] T. Akutsu, Miyano, S., Kuhara, S., "Algorithms for Inferring Qualitative Models of Biological Networks," presented at Pacific Symposium on Biocomputing 5, Hawaii, 2000.

[5] H. Matsuno, Doi, A., Nagasaki, M. and Miyano, S., "Hybrid Petri Net Representation of Gene Regulatory Network," presented at Pacific Symposium on Biocomputing 5, Hawaii, 2000.

[6] H. Alla and R. David, "Continuous and Hybrid Petri Nets," *Journal of Circuits, Systems, and Computers*, vol. 8, pp. 159-188, 1998.

[7] W. Reisig and G. Rozenberg, *Lectures on Petri Nets I: Basic Models*. Berlin: Springer, 1998.
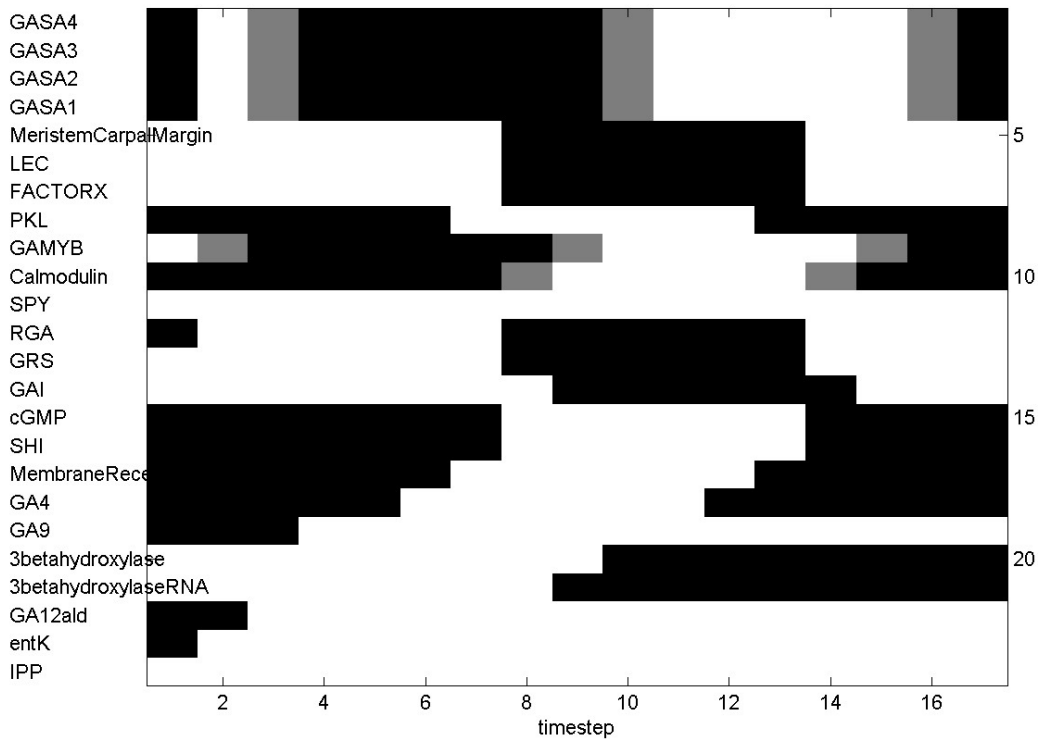
Figure 3. Time history of the sample network for GA metabolism. White means that the node was on and black means that the node was off. Note how the active form of gibberellin (GA4) self-regulates its levels and regulates other downstream processes such as the expression of the GA_MYB trancription factor genes that are inhibited by other transcription factors (RSD, RGA, GAI, SHI).

[8]   V. Hatzimanikatis and K. H. Lee, "Dynamical Analysis of Gene Networks Requires Both mRNA and Protein Expression Information," *Metabolic Engineering*, vol. 1, pp. 275-281, 1999.

[9]   J. A. Dickerson and B. Kosko, "Virtual Worlds as Fuzzy Cognitive Maps," *Presence*, vol. 3, pp. 173-189, 1994.

[10]   B. Kosko, "Fuzzy Knowledge Combination," *International Journal of Intelligent Systems*, vol. 1, pp. 293-320, 1986.

[11]   B. Kosko, "Fuzzy Cognitive Maps," *International Journal Man-Machine Studies*, vol. 24, pp. 65-75, 1986.

[12]   G. A. Banini and R. A. Bearman, "Application of fuzzy cognitive maps to factors affecting slurry rheology," *International Journal of Mineral Processing*, vol. 52, pp. 233-244, 1998.

[13]   K. Lee, S. Kim, and M. Sakawa, "On-line fault diagnosis by using fuzzy cognitive maps," *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences*, vol. E79-A, pp. 921-922, 1996.

[14]   J. A. Dickerson, D. Berleant, Z. Cox, W. Qi, and E. Wurtele, "Creating Metabolic Network Models using Text Mining and Expert Knowledge," presented at Atlantic Symposium on Molecular Biology and Genome Information Systems and Technology (CBGIST 2001), Durham, North Carolina, 2001.

[15]   T. Sun Gibberellin Signal Transduction. Curr. Opin. Plant Biol. vol. 3, pp 374-380

[16]   P. Hedden, A.L. Phillips Gibberellin metabolism: new insights revealed by the genes. Trends Plant Sci. 2000 vol. 5, pp. 523-530.
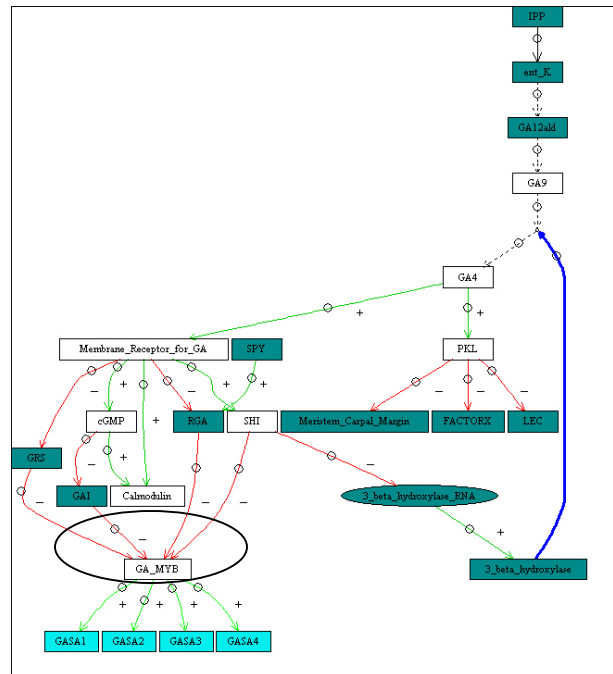
Figure 4: Output of the network at the third time step. Note how GA_MYB is inhibited by other nodes and that the bold catalytic link is activated but does not cause GA4 to be produced.
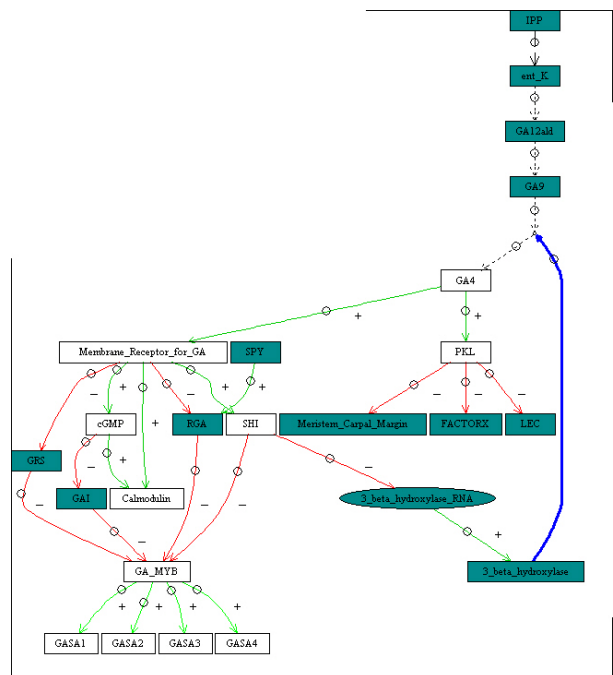

Figure 5: Output of the network at the fourth time step. Note how the catalytic link is activated and GA9 is present so that GA4 can be produced.
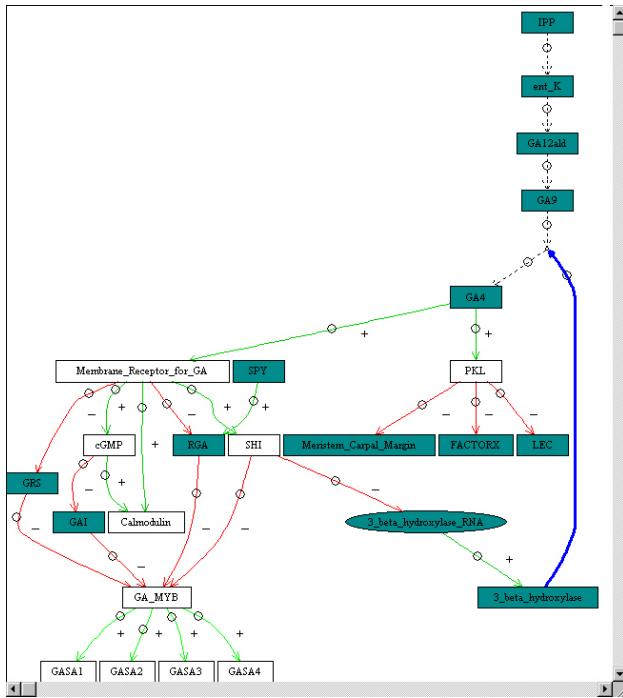
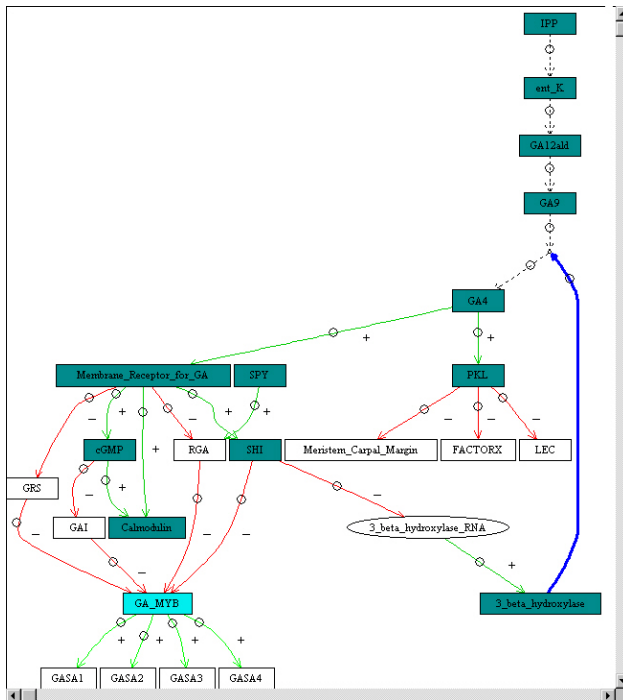Figure 6: Output of the network at the sixth time step. Note how the GA4 node is now active.

Figure 7: Output of the network at the ninth time step. GA4 is negatively regulating the small molecules that inhibit GA_MYB.