Computational Modeling of Biological Pathways

By

Liu Bing

Department of Computer Science
School of Computing
National University of Singapore

2005/06
Honours Year Project Report

Computational Modeling of Biological Pathways

By

Liu Bing

Department of Computer Science
School of Computing
National University of Singapore

2005/06

Project No: H096120
Advisor: Assistant Prof. David Hsu and Prof. P.S. Thiagarajan
Deliverables:
  Report: 1 Volume
Abstract

Computational modeling and analysis of biological networks is important for understanding cellular function at the systems level. It has been widely accepted that there are three main classes of biological networks, namely, metabolic pathways, signaling pathways, and gene regulatory networks (GRN). Hybrid Functional Petri Nets and Bayesian Networks have been proposed to model biological networks, but, those techniques focus on only one particular class of biological networks. Since the signaling pathways and the GRN can interact with each other, we extend the current models to a Hybrid Modeling Framework which links signaling pathways and GRN together. By allowing interactions between them, this framework not only enables us to understand how the GRN responds to external signals but also allows microarray data to be used in validation. We show how to construct a model for the Wnt pathway under this framework. We also validate this model and provide analyses based on the simulation results.

Subject Descriptors:
J.3 Life and Medical Sciences
I.6 Simulation and Modeling
G.1 Numerical Analysis
G.3 Probability and Statistics

Keywords:
Biopathway modeling, Hybrid modeling framework, Hybrid functional Petri nets, Bayesian networks, Signaling pathway, Gene regulatory network

Implementation Software and Hardware:
IBM PC, Windows XP, JDK 1.5.0
Acknowledgement

First of all, I would like to thank my advisors Prof. David Hsu and Prof. P.S. Thiagarajan for their guidance, patience and support during the past year. Without their kind help, I would not be able to complete this project. They introduced me the field of Computational Systems Biology, and gave me a direction for my research. This cutting-edge research is very exciting to me, and working with them is very unforgettable experience, since not only had they taught me how to do research but also I learnt how to be a good researcher. I am very happy to get the chance to still work with them for my graduate studies as a NGS PhD candidate.

I am also grateful to other members of Computational Systems Biology Research Group who help and encourage me. Thanks to my senior Geoffrey Koh. His valuable advice and constant support help me to finish up my project. Special thanks to Lisa Tucker-Kellogg for her very kind help on amending the English problems in my report. I also thank Dr. Manjiri Manohar BAKRE for her concerns on my project, and Leong Waikay for sharing the feeling on this period.

Finally, I am forever indebted to my parents and my girlfriend Han Zheng for their love and encouragement.
## List of Figures

2.1 A Petri net example .................................................. 5  
2.2 HFPN notations ................................................... 6  
2.3 A HFPN example .................................................. 8  
2.4 A BN example ................................................... 9  
3.1 Schematics of the Hybrid Modeling Framework .................. 12  
3.2 Methodology .................................................... 13  
3.3 Mass-action kinetics ........................................... 14  
3.4 Enzyme kinetics ............................................... 15  
3.5 Generalised Noisy-Or model .................................. 18  
3.6 A pathway-GRN interaction example .......................... 22  
4.1 The sketch map of the Wnt/β-catenin pathway .................. 26  
4.2 The HFPN model of the Wnt/β-catenin pathway ............... 28  
4.3 The sketch map of the Wnt/Ca^{2+} pathway ..................... 29  
4.4 The HFPN model of Wnt/Ca^{2+} pathway ....................... 30  
4.5 Crosstalk between two pathways ............................... 31  
4.6 The underlying GRN of Wnt pathways .......................... 32  
4.7 Hybrid modeling the Wnt signaling pathway and its GRN ...... 34  
5.1 Protein concentration time courses I ............................ 38  
5.2 Protein concentration time courses II .......................... 39  
5.3 The comparision of β-catenin concentration ................... 40  
5.4 MYC gene expression ........................................... 40  
5.5 Protein concentration time courses with the effects of feedback .............................. 41
Table of Contents

Title
Abstract
Acknowledgement
List of Figures

1 Introduction
  1.1 Motivation .......................... 1
  1.2 Contributions .................................. 2
  1.3 Organization ................................. 3

2 Background
  2.1 Hybrid Functional Petri nets ......................... 4
  2.2 Bayesian network ................................... 8

3 Modeling
  3.1 Hybrid Modeling Framework .......................... 11
  3.2 Methodology ..................................... 12

4 Implementation
  4.1 Wnt pathways .................................... 25
  4.2 Gene Regulatory network ............................ 31

5 Simulation
  5.1 Simulation engine ................................ 35
  5.2 Simulation results and analysis ...................... 37

6 Conclusions
  6.1 Summary ..................................... 42
  6.2 Limitations and Future Work ........................ 42

References ........................................ 44

A Reactions and Parameters .......................... A-1
Chapter 1

Introduction

1.1 Motivation

Traditional biology has focused on identifying individual components of cell such as life’s most important molecules, genes and proteins, and studying their specific functions. This approach to study biology and human health has left us with a limited understanding of how the human body operates, and how we can best predict, prevent, or remedy potential health problems. As scientists have developed the tools and technologies which allow them to delve deeper into the foundations of biological activity. They have learned that these components almost never work alone, they interact with each other and with other molecules in highly structured but incredibly complex ways, similar to the complex interactions among the countless computers on the Internet.

As an emergent field, Systems biology seeks to understand these complex interactions, as these are the keys to understanding life (Kitanon, 2002). Instead of analyzing individual components, Systems biology focuses on how those components interact with each other to form complex, inter-dependent networks which are called Bio-pathways. Although systems are composed of matter, the essence of a system lies in its dynamics, which cannot be described merely by enumerating components of the system. At the same time, it is misleading to believe that only system structure, such as network topology, is important without paying sufficient attention to the diversities and functionalities of the components. Both structure and components play indispensable roles forming the symbiotic state of the system as a whole.

To understand the dynamics of the system and do quantitative and qualitative analysis,
we need to develop computational models for representing and simulating bio-pathways. In general, there are three classes of bio-pathways: signaling pathways, gene regulatory networks, and metabolic pathways (Cho & Wolkenhauer, 2005). Signaling pathways focus on signal transduction, for example how some stimulation outside the cell membrane can be received and can influence particular proteins and finally affect gene transcription. Gene regulatory networks (GRN) focus on how each gene influences transcription of other genes. GRN models are usually built from microarray data using machine learning for reverse engineering. A metabolic pathway is a series of chemical reactions occurring within a cell, catalyzed by enzymes, to achieve either the formation of a metabolic product to be used or stored by the cell, or the initiation of another metabolic pathway.

Although there is currently a lot of research work being carried out in various research and academic institutions, resulting in a variety of modeling techniques, each has advantages and disadvantages. In this project, we try to bridge some of the gaps in the available models and develop a unified model which allows interconnected signaling pathways and gene regulatory networks. In this model, multiple related pathways can drive the gene regulatory network formed by target genes, which may also regulate proteins in pathways through feedback loops. So far, the only previous work that has flavors of such a connection is cis-reg GRN (Longabaugh, Davidson, & Bolouri, 2005). However, that method is more of a visualization and knowledge management tool rather than a simulation and prediction software.

There are several challenges for connecting such as different representation, time scale, feedback issue, and role of other signaling pathways. For example, a signaling pathway uses concentration unit nM and its time scale is in seconds, while the GRN may be represented by gene expression levels with the time scale in hours. In other hand, some of the signal response genes can produce proteins which can influence pathway, then how to deal with feedback is another important problem to be solved. In this project, we try to overcome those challenges by modeling the interaction between pathways and the GRN.

1.2 Contributions

Our contributions can be summarized as following:

- We extend the current models to a Hybrid Modeling Framework. Under this framework,
a biological network is formed by several signaling pathways and their underlying gene regulatory network. The signaling pathways are studied in mechanical manner and modeled using Hybrid Functional Petri Nets, while the GRN are constructed in inferential manner and modeled using Bayesian Networks. This is the first method to link signaling pathways and GRNs together. By allowing interactions between them, it will enable us to understand how the GRN responds to external signals and also allow microarray data to be used in validation of pathways.

- We demonstrate the idea of our framework by modeling the well-studied Wnt signaling pathway and its GRN, which plays an important role in both disease and development (Veeman, Axelrod, & Moon, 2005). The topology of this model is shown in Figure 4.7 and the parameters used are shown in the appendix.

- We implement a simulation engine and provide several analysis based on the simulation results (including protein concentration time series and gene expression levels) of our Hybrid Modeling Framework model for Wnt, which presented in section 5.2.

1.3 Organization

The rest of the report is organized as follows. Chapter 2 presents some background information. The modeling methodology are described in Chapter 3. After introducing the implementation of the Wnt model in Chapter 4, we provide the simulation results and analyses in Chapter 5. Finally, we discuss future work and conclude in Chapter 6.
Chapter 2

Background

To model biological systems requires models to allow both quantitative and qualitative analyses. In other words, a biological model should not only have a graphical representation to intuitively depict the interaction among components, but also have a mathematical representation to describe the dynamics of the systems. In this chapter, we will provide a brief introduction to the two models used in our Hybrid modeling framework.

2.1 Hybrid Functional Petri nets

The Petri net (PN), originally proposed by Carl Adam Petri in 1962 (Petri, 1962), is a mathematical model for representation and analysis of concurrent, asynchronous, distributed, parallel or stochastic processes. As a modeling language, it graphically depicts the structure of a concurrent system as a directed bipartite graph with annotations. A PN consists of three primitive elements - places, transitions and directed arcs. During modeling, places denote passive entities of the real world such as resources, channels, buffers, or states of a system while active entities such as reactions, operations, or events are represented by transitions. The places are then connected to the transitions (and vice versa) via directed arcs to form a network.

In the graphical representation, places are drawn as circles, transitions are denoted by bars or boxes, and arcs are labeled with their weights (positive integers), where a \( k \)-weighted can be interpreted as the set of \( k \) parallel arcs. Arcs exist only between places and transitions. The input places of a transition are the places from which an arc runs to it; its output places are those to which an arc runs from it.
In modeling, using the concept of conditions and events, the input places and output places of a transition represent the pre-conditions and post-conditions of the event, respectively. Places may contain any nonnegative number of tokens, which are represented as block dots inside the corresponding place. A distribution of tokens over the places of a net is called a marking. Transitions can fire (i.e. execute) if they are enabled, which means there are enough tokens in every input place. When a transition fires, it consumes a number of tokens from each of its input places, and produces a numbers of tokens on each of its output places. A formal definition of a Petri net is given as following (Tadao, 1989):

**Definition 2.1.1** A Petri net is a 5-tuple, \( \mathcal{PN} = (P, T, F, W, M_0) \) where:

- \( P = \{p_1, p_2, \ldots, p_m\} \) is a finite set of places
- \( T = \{t_1, t_2, \ldots, t_n\} \) is a finite set of transitions such that \( P \cap T = \emptyset \)
- \( F \subseteq (P \times T) \cup (T \times P) \) is a set of arcs
- \( W : F \rightarrow \mathbb{Z}^+ = \{1, 2, \ldots\} \) is a weight function
- \( M_0 : P \rightarrow \mathbb{N} = \{0, 1, 2, \ldots\} \) is the initial marking.

![Figure 2.1: A Petri net example](image)

The PN example shown in Figure 2.1 is a simple model of a communication protocol between two processes. In this example, \( t_1 \) is enabled and \( t_2 \) is not enabled. In the next state, if \( t_1 \) fires,
one token will be removed from $p_1$ and one token will be added to $p_2$ and $p_3$. At that state, $t_2$ becomes enabled.

To fill the needs of different application domains, later on, various types of Petri nets have been proposed by extending the original Petri net, such as the timed Petri net, the stochastic Petri net, the hybrid Petri net, and the functional Petri net (Reisig & Rozenberg, 1998). As one of them, the Hybrid Functional Petri net (HFPN) (Matsuno, Tanaka, Aoshima, Doi, Matsui, & Miyano, 2003) inherits the notations of the hybrid Petri net (David & Alla, 1987) and the functional Petri net (Valk, 1978) and adds more semantics and functionality. The terms “Hybrid”, a HFPN can deal with both discrete and continuous components. Thus two kinds of places and transitions are used (the graphical notation are shown in Figure 2.2).

![HFPN notations](image)

Figure 2.2: HFPN notations

A discrete place is the same as a place in PN, i.e. it can only hold integer number of tokens. In other hand, a continuous place can hold non-negative real numbers as its content. For transitions, a discrete transition can only fire when its firing conditions are satisfied for certain duration of time, denoted by a delay function. In contrast, a continuous transition fires continuously in and its firing speed is given as a firing function of values at particular places in the model. The firing speed describes the consumption rate of its input places and the production rate of its output places.

In addition, HFPN also introduce two additional kinds of arcs - the inhibitory arc and the test arc (Figure 2.2). An inhibitory arc with weight $r$ enables the transition to fire only if the content of the place at the source of the arc is less than or equal to $r$. A test arc, behaves like a normal arc, except that it does not consume any content of the place at the source of the arc when it fires. Furthermore, there are also some restrictions for connection. For example, a
discrete place cannot connect to a discrete place via a continuous transition. Test and inhibitory arcs are restricted to only connect incoming places to transitions as they both involve satisfying a precondition. Here is the formal definition (Matsuno et al., 2003):

**Definition 2.1.2**  A Hybrid Functional Petri net $HFPN = (P, T, h, F, a, C, I, O, D, M_0)$ where:

- $P = p_1, p_2, \ldots, p_m$ is a finite set of places
- $T = t_1, t_2, \ldots, t_n$ is a finite set of transitions such that $P \cap T = \emptyset$
- $h : P \cup T \rightarrow \{d, c\}$ is the hybrid function. $p \in P$ is a discrete (resp., continuous) place if and only if $h(P) = d$ (resp., $c$); and $t \in T$ is a discrete (resp., continuous) place if and only if $h(T) = d$ (resp., $c$)
- $F \subseteq (P \times T) \cup (T \times P)$ is a set of arcs
- $a : F \rightarrow \{\text{discrete input arc, continuous input arc, test input arc, discrete output arc, continuous output arc}\}$ is the arc type function.
- $C : T \rightarrow c(m_1(t), m_2(t), \ldots, m_p(t))$, where $m_i(t)$ is the content of source place $P_i$, and $c$ is the firing condition function
- $I : T \rightarrow f(m_1(t), m_2(t), \ldots, m_p(t))$, where $m_i(t)$ is the content of source place $P_i$, and $f$ is the firing speed function. Let $A$ and $B$ be the input and output arc set. If $a(a_i) = \text{test input arc}$, then $f \equiv 0$
- $O : T \rightarrow g(m_1(t), m_2(t), \ldots, m_p(t))$, where $m_i(t)$ is the content of target place $P_i$, and $g$ is the production speed function
- $D : T_{\text{discrete}} \rightarrow d(m_1(t), m_2(t), \ldots, m_p(t))$, where $m_i(t)$ is the content of source place $P_i$, and $g$ is the delay function
- $M_0 : P \rightarrow \mathbb{R}^+$ is the initial marking

A example is shown in Figure 2.3. $t_3$ is a continuous transition, with firing speed function and production speed function $I(t_3) = 0$ and $O(p_3) = 0.1m_1m_4$. Since the firing condition is $m_2 > 3$, i.e. the $p_2$ will inhibit $t_3$ if $m_2 > 3$, thus $t_3$ will be active continuously, producing $0.1m_1 \ast m_4$ content for $p_4$ without consuming $p_2$ and $p_3$ until $m_2 > 3$. Furthermore, $t_2$ is a
discrete transition with delay function $D(p_1) = 2$ and $I(p_1) = O(p_1) = 1$. Thus it will transfer 1 token from $m_2$ to $m_3$ every 2 time unit.

![Figure 2.3: A HFPN example](image)

### 2.2 Bayesian network

A Bayesian network (BN) is a structured directed graph representation of relationships between variables. The nodes represent the random variables, and the edges represent the influence of one variable on another. Its graphical nature is convenient as it facilitates the users to visualize the problem, and in addition it provides a useful method for simplifying the assessment of the joint distribution of a set of random variables by exploiting conditional independencies. Therefore, BNs have been extensively applied in many fields including medicine (Chevrolat, Golmard, Ammar, Jouvent, & Boisvieux, 1998) and computer science (Heckerman, Breese, & Rommelse, 1994).

The following definition uses the style of (Pe’er, 2005). $\mathcal{X} = \{X_1, X_2, \ldots, X_n\}$ is a finite set of random variables, where each variable $X_i$ may take on a value $x_i$ from the domain $Val(X_i)$. We use capital letters such as $X$, $Y$, $Z$ to denote variable names and lowercase letters such as $x$, $y$, $z$ for the values of variables. Sets of variables (resp., values) are denoted by boldface capitals (resp., lowercase) letters such as $\mathbf{X}$ (resp., $x$). In addition, we denote conditional independence as $\perp$. For example, $(X \perp Y|Z)$ means $X$ is conditionally independent of $Y$ given $Z$, i.e. $P(X|Y,Z) = P(X|Z)$. Then we can give a formal definition of a Bayesian network as
Definition 2.2 A Bayesian network is a triple $\mathbf{BN} = \{\mathbf{X}, \mathbf{E}, \mathbf{\Theta}\}$ where:

- $\mathbf{X}$ is a set of random variables
- $\mathbf{E} \subseteq (\mathbf{X} \times \mathbf{X})$ is a set of arcs, and we denote an arc from $X$ to $Y$ as $X \rightarrow Y$, where $X, Y \in \mathbf{X}$
- $\mathbf{\Theta}$ is a set of conditional probability distributions $P(X_i|\text{Pa}(X_i))$, where $\text{Pa}(X_i)$ is the set of parents of $X_i$ in $\mathbf{X}$

An important property of a BN is that it encodes conditional independencies between variables with a graph structure using the Markov Assumption, i.e., every $X_i$ is independent of its non-descendants. Thus, by applying the chain rule of probabilities and the properties of conditional independencies, any joint distribution can be simplified from equation 2.1 to equation 2.2.

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|X_1, X_2, \ldots, X_{i-1})$$ (2.1)

$$P(X_1, X_2, \ldots, X_n) = \prod_{i=1}^{n} P(X_i|\text{Pa}(X_i))$$ (2.2)

This is the so-called chain rule for BN, which can reduce the number of parameters and make the BN more compact. This reduction in parameters is critical when estimating a model from limited data. Robust estimation of a model with many parameters requires many more data points than does estimation of a model comprising fewer parameters.

![Figure 2.4: A BN example](image-url)

As an example, Figure 2.4 shows a BN of five genes. Assume that A is a regulator of C. If A activates C, we expect that in most cases when A is active, so is C. This dependency is gene
expression level is drawn as arc $A \rightarrow C$. Thus, knowing the value of $A$ provides information that can help predict the value of $C$. In addition, $C$ regulates $D$ and $E$; thus, the expression level of $D$ and $E$ is correlated with $C$ and $A$. However, the effect of $A$ on $D$ or $E$ is mediated through $C$, which means that $A$ and $D$ or $E$ are conditionally independent given $C$. Similarly, this network structure also implies: $(A \perp B)$, $(D \perp E|A, B)$, etc. Furthermore, the expression level of $C$ is regulated by both $A$ and $B$. Assuming that $A$ and $B$ each weakly up-regulate $C$, and together they strongly up-regulate $C$, we can represent the probability distribution $\theta_C$ using a conditional probability table (Table 2.1).

<table>
<thead>
<tr>
<th>$a$</th>
<th>$b$</th>
<th>$P(c = 0)$</th>
<th>$P(c = 1)$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0</td>
<td>0.95</td>
<td>0.04</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>0.62</td>
<td>0.38</td>
</tr>
<tr>
<td>1</td>
<td>0</td>
<td>0.73</td>
<td>0.27</td>
</tr>
<tr>
<td>1</td>
<td>1</td>
<td>0.02</td>
<td>0.98</td>
</tr>
</tbody>
</table>

Table 2.1: CPT example

conditioned on $\text{Pa}(X_i) = pa$, where $pa$ is the discrete expression level. Therefore, by the chain rule of BN,

$$P(A, B, C, D, E) = P(E|C, B)P(D|C)P(C|B, A)P(B)P(A)$$  \hspace{1cm} (2.3)

which only requires $1 + 1 + 4 + 2 + 2 = 10$ parameters instead of $1 + 2 + 4 + 8 + 16 = 31$ parameters for the case without any independence assumptions,

$$P(A, B, C, D, E) = P(E|D, C, B, A)P(D|C, B, A)P(C|B, A)P(B|A)P(A)$$  \hspace{1cm} (2.4)
Chapter 3

Modeling

This chapter presents the details of our modeling techniques. We begin with an overview of the *Hybrid Modeling Framework*, and then describe the modeling methodology for the pathway, the GRN and their interactions. After that, we will address the parameter estimation issues.

3.1 Hybrid Modeling Framework

To solve the problems mentioned in Chapter 1, we propose a hybrid modeling framework which links signaling pathways and gene regulatory network together. Figure 3.1 shows a graphical representation of our framework. Within this framework, a biological network is formed by multiple signaling pathways and their one underlying gene regulatory network. The signaling pathways are studied in a mechanical manner (i.e. using interaction and dynamical time course measurements such as Western Blot data) and modeled using HFPNs, while the gene regulatory network is constructed in an inferential manner (i.e. using high-throughput multi-omic measurements such as microarray data) and modeled using a BN. This hybrid framework allows such interactions, not only between the signaling pathways and gene regulatory network, but also amongst themselves. In this way, it will enable us to understand how the gene regulatory network responds to external signals and also allow microarray data to be used in validation of pathways.

In summary, the word "hybrid" in our framework has three meanings referring to three issues we have integrated:

- Mechanistic and inferential sources of data
Continuous and discrete time scales
Quantitative and qualitative analyses

The details of these meanings will be explained in the later sections.

Figure 3.1: Schematics of the Hybrid Modeling Framework

3.2 Methodology

The methodology for the hybrid modeling we proposed is diagrammed in Figure 3.2. It includes three phases: model construction, parameter estimation, and experimental validation. Model construction can also be divided into three steps: modeling the pathway, modeling the GRN, modeling their interactions. Once the model has been created, we can use experimental data to estimated unknown parameters in the model. Then the model can generate specific predictions that can be tested experimentally. Furthermore, experimental validation can be used to iteratively refine the model and generate new predictions.
3.2.1 Modeling pathway

For the first step of model construction, we model the pathway as follows: we extract the skeleton network of signaling pathways are extracted from the biological literature (e.g. Entrez-PubMed) and by mining public databases (e.g. KEGG). Then we represent the various interactions in the network using HFPN. Here we divide a cell system into three compartments: extracellular space, cytoplasm and nucleus. One of the main assumptions of biochemical modeling is that molecules are evenly distributed throughout the entire system compartment, such that their concentrations can be represented by single variables. In the HFPN model, this would be represented by a continuous place whose value denotes the concentration of a particular protein type. The presence or the absence of certain conditions, such as availability of serum, can be modeled using discrete places instead. Since there are different types of interactions that can occur between proteins, we use different kinetic models to represent them.

Mass-action kinetics

The mass-action law states that the rate of any elementary reaction is proportional to the concentration of the molecules involved, irrespective of the identity of the molecules. So in a signaling pathway, all reactions that do not chemically modify their proteins will follow the mass-action law, including association, dissociation and translocation. Association is the binding
process of two proteins, while the opposite process is called dissociation. Translocation involves the molecules moving from one compartment of the cell to another, such as from the cytoplasm to the nucleus. Figure 3.3 shows the equations denoting those three processes and their HFPN equivalent.

Figure 3.3: Mass-action kinetics

Enzyme kinetics

In the living cell, many reactions are catalyzed by enzymes and follow enzyme kinetics. The following equation shows a possible reaction scheme of such a reaction:

\[ P + E \rightleftharpoons P.E \rightleftharpoons P.S \rightarrow P + S \]

where \( S \) is the substrate, \( P \) is the product and \( E \) denotes the enzyme that catalyzes the reaction and which will not be consumed. Under the assumption that enzyme concentrations are in general much lower than the concentrations of substrates and products, we have decided to adopt the Michaelis-Menten equation (Michaelis & Menten, 1913) for modeling enzyme-mediated protein modification in the pathways. Michaelis-Menten model is based on the quasi steady state approximation for chemical reactions. Figure 3.4 shows the HFPN models representing two types of enzyme catalyzed reactions under this scheme. In the first model, the arc from the place representing the enzyme \( E \) is a test arc, as no enzyme molecules are consumed in a catalytic reaction. In the rate law equation, the parameters \( k \) is the catalytic constant and \( K_M \) is the Michaelis constant, which indicates the affinity of the substrate to the enzyme (where low values indicate that the \( ES \) complex is held together very tightly). In cases where we know that the reaction is an enzyme-catalyzed reaction, but not the enzyme involved, the rate law

14
equation can be written in the form depicted by Panel B of Figure X. In this form, \( V_{max} \) is the maximal velocity of the reaction, and is an experimentally derived value.

\[
v = \frac{k[S][E]}{(K_m + [S])}
\]

\[
v = \frac{V_{max}[S]}{(K_m + [S])}
\]

Figure 3.4: Enzyme kinetics

There may be other types of chemical reactions in the living cel, but since we will concentrate on protein interactions, these two types of kinetics (mass action and enzymatic) will be sufficient for the pathway that follows. In the next chapter, we will see how to use them to systematically convert the pathway into its HFPN equivalent in our Wnt model.

### 3.2.2 Modeling GRN

The reason we adopt BN as the tool for modeling GRN is that it not only provides a way to account for the noise and data limitations inherent in expression studies, but also provides a way to retain the combinatorial logic of transcription regulation. In the BN model, each vertex \( v_i \in V \) denotes a gene, and the random variable \( X_i \) describes the normalized discrete expression level of its respective gene. The regulatory influences are described by the arcs among vertices. In other words, the regulators of a particular gene are denoted as its parents in the network. Therefore, for each \( X_i \), a conditional probability \( Pr(X_i|Pa(X_i)) \) is defined to quantify the regulatory influence of \( Pa(X_i) \) on \( X_i \). We discretize gene expression levels into trinary values: 0, 1 and 2, i.e. \( X_i = 0 \) means gene \( x_i \) has been down-regulated, \( X_i = 1 \) means gene \( x_i \) keeps normal expression and \( X_i = 2 \) means gene \( x_i \) has been up-regulated. Figure 2.4 in the previous chapter gives an example of a small GRN consisting of 5 genes. To follow the
Markov assumption, we should assume that the probability of each gene depends only on the state of its regulators, and then the joint probability can be expressed as equation 2.3. Thus, for example, the CPT of gene B is shown in Table 3. With those CPTs, we can finally compute probabilities of any gene \( X \) being up-regulated, down-regulated or unchange. For example:

\[
\]

<table>
<thead>
<tr>
<th>( A )</th>
<th>( P(B = 0) )</th>
<th>( P(B = 1) )</th>
<th>( P(B = 2) )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.8</td>
<td>0.1</td>
<td>0.1</td>
</tr>
<tr>
<td>1</td>
<td>0.3</td>
<td>0.4</td>
<td>0.3</td>
</tr>
<tr>
<td>2</td>
<td>0.0</td>
<td>0.1</td>
<td>0.9</td>
</tr>
</tbody>
</table>

Table 3.1: CPT example

Construction of the GRN using expression data

Humans have more than 30,000 genes and it is very difficult identify all the interactions between them. So, to determine the topology of our GRN, we can adopt the approach proposed by Friedman which constructs BN based on multiple expression measurements such as microarray data (Friedman, Linial, Nachman, & Pe’er, 2000). Given a training set \( T \) of independent realizations of the \( n \) random variables \( X_1, X_2, \ldots, X_n \), the problem is to find a BN that best matches \( T \). With their approach, the network structure is usually determined using a heuristic search approach or Markov Chain Monte Carlo method. For each network structure visited in the search, an algorithm learns the maximum likelihood parameters for the conditional distribution functions. It then computes a score for each network using the Bayes Information Criteria (Neapolitan, 2004) or some other metric that measures the overall fit of the model to the data. A common solution is to assign a score to each calculated network using the a posterior probability of the calculated network, \( N \), given the training data by

\[
logP(N|T) = logP(T|N)P(N)/(P(T) = logP(T|N) + logP(N) + \text{const}
\]

where the constant is independent of the calculated network and \( P(T|N) = \int P(T|N, \Theta)P(\Theta|N)d\Theta \) is the marginal likelihood that averages the probability of the data overall possible parameter as-
signments to the network. The choice of the a priori probabilities $P(N)$ and $P(\theta|N)$ determines the exact score. One may then select the highest-scoring network as the best network.

During the execution of the BN learning algorithm, one of the most important challenges is the number of parameters. As we mentioned in chapter 2, given a network (consisting of $n$ nodes with indegree $k_i$ for each node $i$), taking the conditional independencies into account can reduce the total number of parameters from $(p - 1)(p^n - 1)$ to $\sum (p - 1)p^{k_i}n$, where $p$ is the number of states of each node. However, we still have to deal with large numbers of parameters if $p$ and $k$ are relatively big. For example, when $n = 50$, $p = 5$, $k \approx 3$, there will be nearly 62500 parameters! Even the tiny GRN shown in Figure 2.4 requires $2 \times (1 + 3 + 9 + 3 + 1) = 34$ parameters. Therefore, we suggest labeling each arc with a link strength value $\mu$ so that the CPT can be derived from it. In this way, the number of parameters required will exactly equal the number of arcs: $\sum k_i$. For example, the network shown in Figure 2.4 only needs 4 parameters. This reduction in parameters is critical when estimating a model from experimental data.

**Link strength**

Kokkonen et al. have presented a method specifying how the CPT can be described with the aid of link strength parameters that are assigned to each link from a parent node to a child node, and that attain values from -1 to 1 (Kokkonen, Koivusalo, Laine, Jolma, & Varis, 2005). Their method combines the concept of the link strength, introduced by (Varis & Kuikka, 1994) and the *generalized Noisy-Or model* of Srinivas (Srinivas, 1993).

The *original Noisy-Or model* considers the case where any one of a set of conditions is likely to cause the same effect in the child node, and it presumes each node to be Boolean. The *generalized Noisy-Or model* extends the original by allowing nodes to take on an infinite number of possible states. Figure 3.5 shows the schematic of the *generalized Noisy-Or model*. The state of each parent node $u_i$ is passed through a line failure device $N_i$ that yields as an output $u'_i$. The function $F$ maps the output from all failure devices, $u' = [u'_1, u'_2, \ldots, u'_n]$, to the state of the child node $x$ where $n$ is the number of parent nodes. The aim is to create the CPT $P(x|u)$ relating the states of the parent nodes $u$ and the states of the child node $x$ with aid of the failure devices $N_i$ and the mapping function $F$.

$$P(x|u) = \sum_{u':x=F(u')} P(u'|u) = \sum_{u':x=F(u')} \prod_{u'} P(u'_i|u_i)$$
To parameterize the conditional probabilities \( P(u'|u) \), line failure probabilities need to be determined for each state in each parent node. The link strength approach of Varis and Kuikka provides a way to describe a CPT relating two nodes with the aid of a single link strength parameter. Kokkonen modified their method as follows:

\[
P(u'_i = r | u_i = c) = \begin{cases} 
\frac{1}{m_i} + |\eta_i|K, & r = c \\
\left(|\eta_i| - \frac{(c-r)^2}{2} + \frac{1 - |\eta_i|}{m_i - 1}\right) \left(1 - \frac{1}{m_i} - |\eta_i|K\right), & r \neq c 
\end{cases}
\]

where \( r \) is the index of the state \( u'_i \), \( c \) is the index of the state \( u_i \), \( \eta_i \) is the link strength of the \( i \)th node, and \( K = 1 - 1/m_i \). The \( F \) function was written in the following form:

\[
F(u') = x \left(\frac{1}{\sum_i |\eta_i|} \sum_i (|\eta_i|H(u'_i))\right)
\]

where

\[
\begin{align*}
H(u'_i) &= I(u'_i), & \eta_i \geq 0 \\
H(u'_i) &= m_i + 1 - I(u'_i), & \eta_i < 0
\end{align*}
\]

Figure 3.5: Generalised Noisy-Or model
of these states and $I(x)$.

$$w_{\text{floor}}(I(x)) = \text{ceiling}(I(x)) - I(x)$$

$$w_{\text{ceiling}}(I(x)) = I(x) - \text{floor}(I(x))$$

For example, if $I(x) = 2.8$ then states $I(x) = 2$ and $I(x) = 3$ are assigned weights 0.2 and 0.8, respectively. In addition, the values of the conditional probabilities $P(x|u)$ have to be adjusted with a correction such that $P(x|u)$ becomes non-informative when all link strengths approach zero. For two parents and three states, the correction $c_2$ is derived from the following equation:

$$c_2 = \frac{|\eta_{m\text{in}}|}{|\eta_{m\text{ax}}|}c_1 + (1 - \frac{|\eta_{m\text{in}}|}{|\eta_{m\text{ax}}|})w$$

$$c_1 = |\eta_{m\text{ax}}|w + (1 - |\eta_{m\text{ax}}|)b_iw$$

Where $w$ is the weight resulting from a non-integer state index, and the values of $b_i$ are based on the distribution of the arithmetic average of uniformly distributed variables. As a result, this approach guarantees the relationship between link strength parameters and the resulting $P(x|u)$. The guarantee is intuitively reasonable, since the following properties can be satisfied:

- Property 1. Inclusion of a parent node with a link strength value of zero should give an identical result to the case where that node is absent.
- Property 2. Equal link strength values should result in an equally strong effect on $P(x|u)$.
- Property 3. The effect of a negative link strength should have the same magnitude as the effect of an equally strong positive link, but be in the opposite direction.
- Property 4. When all link strengths are zero, the CPT should be non-informative, i.e. all probabilities are equal to the inverse of the number of states in $x$.

However, in GRNs, some genes have coupled regulators which can only take effect with the presence of both regulators simultaneously. In this case, the relation between parents and child became an generalized Noisy-And model. Therefore, we extend Kokkonen’s method by modifying the weight equation as follows:

$$w_{\text{floor}}(I(x)) = I(x) - \text{floor}(I(x))$$

$$w_{\text{ceiling}}(I(x)) = \text{ceiling}(I(x)) - I(x)$$
3.2.3 Modeling interaction

From the above, we have seen that the techniques used for modeling signaling pathways and GRNs are quite different. To link them into a unified model, there are many challenges such as time scale, representation and feedback. Firstly, in our pathway model, the time scale is in seconds or minutes meaning that the system will reach the steady state in for example 30 minutes. By contrast, in the GRN model, a BN can only tell us the effect of one gene on another, and it cannot tell how much time that effect will take. In fact, indirect gene regulation can be considered as an abstraction of multiple pathways in the cell. Since it may take effect via a very short pathway or very long one and there are also many uncertain factors, we cannot know the duration of the regulation process. Secondly, in our pathways we are using concentrations of proteins (nM) while in the GRN we are using the probabilities of normalized gene expression levels. Thirdly, we must also be aware that some of the signal response genes can produce particular proteins which can influence the pathway. Then how to deal with feedback from the GRN to the pathways is the most important problem to be solved.

In our proposed model, we employ a novel hybrid timescale to take up the first challenge. In detail, signaling pathways are simulated in continuous time (in minutes) while the GRN are executed using discrete time steps that take 4 hours. Here we assume most of the regulatory process can be finished within 4 hours. Under this assumption, the expression probabilities of genes in the GRN will be computed and updated every 4 hours, and then some of the genes may feedback to the pathway and update the concentration of corresponding proteins. Therefore, the interaction between the pathway and the GRN can be divided into the following two processes - From pathway to GRN and From GRN to pathway, which denoted by yellow and white arrow respectively in Figure 3.1.

From pathway to GRN

After the cell receives external stimulation, the signal will be relayed downstream and may finally get into the nucleus and activate or form particular transcription factors that can influence gene expression. In other words, transcription factors, the end point of a signaling pathway, are the crucial cut vertices of the hybrid network. To quantify the effects of transcription factors, we assume that the concentration of transcription factors is proportional to the link strength of the
corresponding arcs in the GRN. It implies that the more transcription activators (or expressers) exist, the higher the probability of selected genes being up-regulated (or down-regulated). So the set of chosen arc candidates can be defined in this way:

**Definition 3.2.3** Arc candidate set \( A = \{ a \rightarrow b | a \text{ is crucial protein of pathway } P, \ TF \text{ is end-point of } P, b \in \text{ the direct target gene of } TF \} \)

Then for each candidate arc \( a \in A \), the effects of pathway on GRN can be calculated in this way:

\[
\mu_a = k_a |B|
\]

where \( |B| \) is the concentration of the crucial protein in pathway.

For example, in the simple pathway shown in Figure 3.6, after receiving the signal of \( A \), receptor \( B \) can activate protein \( C \) (denoted as \( C^* \)) which will go into the nucleus and become a transcription factor (\( nC^* \)) that up-regulates gene \( x \) and gene \( y \). Under our definition, gene \( c \) which can produce \( C \), is the crucial protein. Since \( nC^* \) is the end-point transcription factor of this pathway with direct target genes \( x, y \), we can select signal \( A \) as \( c \rightarrow x, c \rightarrow y \). Then

\[
\mu_{c \rightarrow x} = k_{c \rightarrow x} |C^*| \text{ and } \mu_{c \rightarrow y} = k_{c \rightarrow y} * |C|
\]

**From GRN to pathway**

As mentioned in chapter 1, in the nucleus, transcription factors often influence a number of genes, and some of these target genes can be translated into proteins which may subsequently influence transcription of other genes. Among those signal response genes, we define feedback genes as those genes which produce proteins that can affect the signaling pathway. For example, in the GRN in Figure 3.6, both gene \( b \) and gene \( y \) are feedback genes. For gene \( b \), protein \( B \) is a component in the pathway. More protein \( B \) will amplify the signal strength, so we say \( b \) is a positive feedback gene. For gene \( y \), protein \( Y \) can deactivate protein \( C \), and we say \( y \) is a negative feedback gene. After identifying the feedback genes in the GRN, we can setup feedback links under the assumption that the protein concentrations depend on the expected value of the corresponding gene. In detail, for each discrete time step (4 hours), the concentration of every feedback gene product will be updated in the following way:

\[
[B] = k_b[B] \sum_{i=0}^{2} P(X_b = i)i, \text{ where protein } B \text{ is the product of gene } b
\]
Figure 3.6: A pathway-GRN interaction example

The biological process behind this approach is that first of all, the concentration of mRNA in the cell will be correspond to the expression level of its gene. Then after the ribosome translates those mRNA molecules into amino-acid chains, the amino-acid chains will fold into particular proteins. Our method abstract this process by assuming the concentration of gene product proteins is proportional to the gene expression levels which are reflected by the expected value of gene expression probabilities.

3.2.4 Parameter estimation

To complete the model, now we come to another important part which is also the greatest bottleneck for quantitative modeling - estimating parameters for the various rate equations. Obviously, for a large network, it is impossible to determine all the parameters by setting up experiments, since it requires huge resource and is extremely time-consuming. Moreover, some parameters may not be measured due to technical difficulties. Thus besides mining kinetic data from the literature and public databases, we also have to estimate parameters based on limited and noisy experimental data such as time course data of specific proteins in response to various stimulations.
Parameter estimation can be stated mathematically as a nonlinear global optimization problem determining the values of parameters set $P$ such that they minimizes the difference between the experimental data and the simulation results. The difference is usually quantitatively measured using a global cost function. Ideally, experimental data would be available under multiple conditions because more conditions provide a more constrained and accurate model. For example, we can grow the cell or biological system with the addition of different drugs, knockout a particular gene or knockdown the RNA level. The following cost function $J(P)$ includes fitting multiple experiments under various conditions:

$$J(P) = \sum_{i,j,t} \sqrt{\frac{(x_{ij}(t_e, P) - x_{ij}^{exp}(t_e))^2}{w_{ij}}}$$

where $x_{ij}(t_e, P)$ is the simulation data point for the $i$th element in experiment $j$ using parameters $P$ at time $t_e$, and the corresponding experimental data points are denoted as $x_{ij}^{exp}(t_e)$. $w_{ij}$ is the weight used for normalization.

Various approaches have been proposed to minimize global cost functions (Moles, Mendes, & Banga, 2003). Naturally, those algorithms require searching through the solution space to find suitable $P$. For example, the Levenberg-Marquardt optimization algorithm uses information on the slope $dJ/dP$ and curvature $d^2J/dP^2$ to find a better fit in the solution space (Marquardt, 1963). However, global optimization is still very difficult since the search space is too large especially for a huge network. To solve this problem, our group is working on reducing the search space by decomposition of the network into smaller components (Koh, Teong, Hsu, Clement, & Thiagarajan, 2005).

The first level of decomposition is based on the property of catalytic reaction. Since we adopt Michaelis-Menten kinetics to model such reactions, the enzymes are not be consumed during reactions. Thus we can decompose the network into separate modules such that the molecules within a module do not flow to other modules. Then the entire network will be simplified into a DAG, where vertices denote modules and arcs denote inhibition or activation. After ranking the vertices in the DAG according to topology and experiment data, we can estimate the parameters of each modules using an optimization algorithm. For better performance, here we chose a variant of the Evolutionary Strategies approach (Beyer & Schewefel, 2002) which is shown in Algorithm 1.
Algorithm 1 \((\mu + \lambda) - ES\) algorithm

1: generate parent vectors \(p_i, i = 1, 2, \ldots, \mu\)
2: repeat
3: create offsprings \(c_l \leftarrow \text{recombination}(p_{\text{rnd}_1}, p_{\text{rnd}_2}), l = 1, 2, \ldots, \lambda\)
4: for all child \(c_l\) do
5: \(c_l \leftarrow \text{mutate}(c_l)\)
6: \(score_l \leftarrow \text{fitness}(c_l)\)
7: end for
8: select the \(\mu\) most fit ones from the \((\mu + \lambda)\) candidates
9: until no better parameters could be obtained

Let us summarize how the pathway and GRN are integrated together under our Hybrid Modeling Framework. The signaling pathway and its interaction with the GRN are modeled in a mechanical manner while the GRN is inferred from expression data using a reverse-engineering method. This explains the first meaning of word “hybrid” - Mechanistic and inference models. For the second meaning - Continuous and Discrete time scales, we have already described in section 3.2.3, discrete time steps are converted to continuous intervals of 4 hours. Because our model can predict not only system dynamics (e.g. time course data of protein) but also signal response genes (e.g. gene expression data), both quantitative and qualitative analyses can be conducted. This is the third meaning of “hybrid” in our modeling framework. In next chapter, we will show how to model an important pathway - the Wnt signaling pathway and its GRN, using our Hybrid Modeling Framework.
Chapter 4

Implementation

In this chapter, we aim to demonstrate the idea of our Hybrid Modeling Framework by modeling the well-studied Wnt signaling pathway and its GRN. In the past 20 years, more than 2000 papers have been published about Wnt signaling as it plays an important role in both disease and development (Veeman et al., 2005). We will first give a brief introduction about Wnt signaling following with HFPN models for pathways. After that we will describe the GRN portion of our Wnt model.

4.1 Wnt pathways

The Wnt proteins are a large family of cysteine-rich secreted ligands that control development in organisms including Caenorhabditis elegans, Drosophila, Xenopus, zebrafish, chicken, mouse, and humans (Dale, 1998). In addition, they are also implicated in the genesis of human cancer (Polakis, 2000). Because the Wnt genes are defined by a sequence homologous to the original members WNT1, first discovered in the mouse that causes mouse mammary tumorigenesis (Nusse & Varmus, 1982), the Wnt proteins are involved in more than one pathway which regulate many stages of development (Nusse, 2003). The most well-studied pathway is known as the Canonical Wnt pathway, which can results in stabilization of the transcriptional co-activator β-catenin (Nusse, 2005). A continuing update on this pathway, including figures and gene tables, can be found on the Wnt homepage http://www.standford.edu/~rnusse/wntwindow.html. Other pathways (β-catenin independent) involving Wnt proteins which can affect cell development and differentiation via other means are called Non-canonical Wnt pathways. Within the past 15
years, at least three non-canonical Wnt pathways have been proposed. One pathway involves activation of calcium/calmodulin-dependent kinase II (CamKII) and protein kinase C (PKC). Another includes recruitment of heterotrimeric GTP-binding proteins to activate phospholipase C (PLC) and phosphodiesterase (PDE). Lastly, a pathway similar to the planar cell polarity (PCP) pathway in Drosophila has been identified that activates the Jun-N-terminal kinase (JNK) and, perhaps, small GTP-binding proteins (Kohn & Moonb, 2005).

As suggested by Arias et al, the future understanding of Wnt signaling may require models that are based on a signaling network rather than a single linear pathway (Arias, Brown, & Brennan, 1999). We attempt to model both the Canonical Wnt pathway (Wnt/β-catenin) and one of the Non-canonical Wnt pathways (Wnt/Ca$^{2+}$ pathway) with their cross-interactions. Hence, the pathway portion of our model consists of two main parts - the Wnt/β-catenin pathway and the Wnt/Ca$^{2+}$ pathway.

4.1.1 Wnt/β-catenin pathway

A simplified version of current model of Wnt signal transduction is presented in Figure 4.1 (Fearon & Cadigan, 2005).

![Figure 4.1: The sketch map of the Wnt/β-catenin pathway](image-url)
The key effecter of the canonical Wnt pathway is β-catenin, the Drosophila version of which is Armadillo. In the absence of the Wnt signal, β-catenin can be phosphorylated by a large complex that involves Axin, adenomatous polyposis coli (APC) protein and glycogen synthase kinase 3β (GSK3β). As a result, the levels of cytosolic β-catenin are normally kept low through continuous proteasome-mediated degradation. Subsequently, the nuclear beta-catenin levels are also kept low. In this case, the transcription co-factors such as lymphoid enhancer-binding factor1/Tcell-specific transcription factor (LEF/TCF) will bind to TLE/Groucho and act as transcription repressors, preventing the expression of Wnt target genes (Daniels & Weis, 2005).

However, when Wnt ligands bind to the co-receptor - Frizzled (Fz) and low density lipoprotein receptor-related protein (LRP) on the cell surface, these receptors can relay the Wnt signal to intracellular proteins in two ways: The Wnt-induced Fz can activate Dishevelled (Dsh) via some unknown mechanism (Cong, Schweizer, & Varmus, 2004) (for simplicity we assume it to be phosphorylation) and the Wnt-induced LRP can recruit Axin to bind to its tail (Mao, Wang, Liu, Pan, Farr, Flynn, Yuan, Takada, Kimelman, Li, & Wu, 2001). Since Dsh may inhibit the binding of Axin/APC and GSK3β, Axin is a member of Axin/APC/GSK3β, both of two signal branches will destruct the complex Axin/APC/GSK3β which is used to phosphorylated beta-catenin. As a consequence, the degradation of beta-catenin is inhibited, and this protein accumulates in the cytoplasm and nucleus. As nuclear β-catenin level is kept high, it then directly displaces TLE/Groucho repressors from LEF/TCF and act is transcription activator, active the expression of target genes (Daniels & Weis, 2005).

In addition, there are some Wnt pathway inhibitors exist in both extracellular space and cytoplasm such as Dickkopf (Dkk) and β-TrCP. Dkk can inhibit Wnt signal by binding with LRP, altering its conformation (Es, Barker, & Clevers, 2003). β-TrCP is part of the SCF Uniquitin ligase complex, adding ubiquitin to phosphorylated beta-catenin, marking it for degradation by Proteasome (Maniatis, 1999).

The HFPN model of the canonical Wnt pathway is shown in Figure 4.2. Part A shows the receptor level and part B and part C depicts the formation of the degradation complex consisting of APC, Axin and GSK3β and the degradation core cycle where β-catenin gets phosphorylated by the degradation complex. Here the reaction scheme is adapted from Lee’s model (Lee, Salic, Kruger, Heinrich, & Kirschner, 2003). At the end of the pathway, part D shows the cytosolic β-catenin being translocated into the nucleus and binding to its co-factor TCF/LEF to start
Figure 4.2: The HFPN model of the Wnt/β-catenin pathway
the transcription of its downstream targets. An additional link was added from part B to part D as APC can export beta-catenin from nucleus, down-regulating expression (Woodgett, 2001).

### 4.1.2 Wnt/Ca$^{2+}$ pathway

Comparing with the Wnt/β-catenin pathway, the Wnt/Ca$^{2+}$ pathway is quite straightforward. The scheme is shown in Figure 4.3. Wnt5a, after binding to Rfz2 (Veeman et al., 2005), will cause the cytosolic G-Protein (GPCRs) to break of into its α and β/gamma components (Wang & Malbon, 2003). The $G_{\alpha}$ subunit will activate phosphodiesterase (PDE), which will inhibit cyclic guanosine monophosphate (cGMP) activity in the cell. The $G_{\beta\gamma}$ subunits will activate phospholipase C (PLC), which in turns hydrolyzes phosphatidylinositol 4,5-bisphosphate (PIP2) into inositol 1,4,5-trisphosphate (IP3) and diacylglycerol (DAG). DAG will activate protein kinase C (PKC) while IP3 increases the concentration of free intracellular calcium (Ca$^{2+}$) (Alberts, Johnson, Lewis, Raff, Roberts, & Walter, 2002), activating calcineurin and CamKII. Finally, calcineurin activates nuclear factor of activated T cells (NF-AT) by dephosphorylating it, increasing its affinity to bind to DNA (Rao, Luo, & Hogan, 1997). Consequently, the transcription factor NF-AT will influence expression of its target genes. Figure 4.4 is the HFPN model for the Wnt/Ca$^{2+}$ pathway.

![Figure 4.3: The sketch map of the Wnt/Ca$^{2+}$ pathway](image)
Figure 4.4: The HFPN model of Wnt/Ca\(^{2+}\) pathway
4.1.3 Crosstalks

Since the interaction among pathways is one of our major concerns, we have identified three possible interactions between the $Wnt/\beta$-catenin pathway and the $Wnt/Ca^{2+}$ pathway. First of all, the two pathways may cross-talk due to the effects PKC has on Dsh, i.e. PKC from the $Wnt/Ca^{2+}$ pathway may inhibit $\beta$-catenin by phosphorylating Dsh, a component of the $Wnt/\beta$-catenin pathway. Secondly, another level of interaction could be CamKII (another component of the $Wnt/Ca^{2+}$ pathway) directly phosphorylating LEF, a transcription co-factor in the canonical pathway. Effects are seen in Xenopus but could also be possible in vertebrates (Kuhl, Geis, Sheldahl, Pukrop, Moon, & Wedlich, 2001). Lastly, inhibitors of cGMP PDEs are reported to inhibit the $Wnt/\beta$-catenin pathway by increasing intracellular concentrations of cGMP, activating $cGMP$-dependent protein kinase (PKG) which may reduce $\beta$-catenin concentration by phosphorylating, and promoting apoptosis (Li, Pamukcu, & Thompson, 2002). To model those interactions, the links shown in Figure 4.5 are added.

![Diagram](image)

Figure 4.5: Crosstalk between two pathways

4.2 Gene Regulatory network

As mentioned in previous chapter, the construction of GRN is base on microarray data using reverse-engineering techniques. However, when the number of genes involved is big, this process may need a large number of data set with various treatments. For example, Gardner has inferred a BN of the E. coli SOS pathway form a set of 56 microarrays measuring cell response to single-transcript perturbations. Due to the time constraint, our collaborator from Genome Institute
of Singapore hasn’t provided enough experiment data. Instead, we construct the GRN of the Wnt model using the physical strategy, identifying gene regulations from literature. We first look for the direct target genes of two pathways. For the Wnt/β-catenin pathway, transcription factor TCF/LEF/beta-catenin can up-regulate many important genes such as MYC, JUN, CCND1, FOSL1, PPARD, MMP7, etc. and down-regulate genes such as Osteocalcin, Dpp, stripe, etc. For the Wnt/Ca^{2+} pathway, the target genes for NF-AT can be divided into those that absolutely require cooperation for activation (e.g. IL2, GM-CSF, IL3, IL4 and FasL), and those that are induced by NF-AT alone (e.g. TNF, IL-13, and Cox2) (Masuda, Naito, Tokumitsu, Campbell, Saito, Hannum, Arai, & Arai, 1995). The products of some of the Wnt target genes are transcription factors (e.g. c-myc, c-jun and fra-1), so we continuously identify the target genes of them. Next we try to find out the additional interaction links among those gene candidates and finally build our GRN shown in Figure 4.6. This GRN consists of more than 50 genes, and every link is supported by a publication. Since there are no cycles in BN, the root nodes are colored as green. Moreover, the important feedback genes are colored as pink. Table 4.1 summarizes the properties of those genes.

![Figure 4.6: The underlying GRN of Wnt pathways](image-url)
Table 4.1: Feedback target genes

<table>
<thead>
<tr>
<th>Gene</th>
<th>Property</th>
<th>Interacts with</th>
<th>Ref (PMID)</th>
</tr>
</thead>
<tbody>
<tr>
<td>DKK</td>
<td>-</td>
<td>LRP</td>
<td>15378020</td>
</tr>
<tr>
<td>FZ</td>
<td>-</td>
<td>Wnt</td>
<td>9876186</td>
</tr>
<tr>
<td>FZ2</td>
<td>+</td>
<td>Wnt</td>
<td>9630221</td>
</tr>
<tr>
<td>AXIN2</td>
<td>-</td>
<td>β−catenin</td>
<td>1175446</td>
</tr>
<tr>
<td>BTCRP</td>
<td>-</td>
<td>β−catenin</td>
<td>10882123</td>
</tr>
<tr>
<td>LEF1</td>
<td>+</td>
<td>β−catenin</td>
<td>11326276</td>
</tr>
<tr>
<td>TCF1</td>
<td>-</td>
<td>TCF</td>
<td>1048937</td>
</tr>
</tbody>
</table>

Thus the Wnt model, an implementation of our modeling methodology, is completed. Figure 4.7 shows the full model which combining two Wnt pathways and GRN. In the following chapter, we will talk about the simulation issues and offer some analyses.
Figure 4.7: Hybrid modeling the Wnt signaling pathway and its GRN
Chapter 5

Simulation

This chapter presents the simulation issues. We will first describe the implementation of our simulation engine. After that we will offer some analyses based on the simulation results of the Wnt model.

5.1 Simulation engine

Although there are tools that can be used to execute HFPN such as Cell Illustrator\textsuperscript{TM} (Doi, Nagasaki, Fujita, Matsuno, & Miyano, 2003) and also some BN modeling tools such as BNJ, none of them can process both. Due to the hybrid feature of our model, we wrote our own simulation engine in Java. A parser can load the model in text format as the input of the system. During loading, the pathway portion will be encoded as HFPN components: places and transitions. Each place $p_i$ maintains two attributes: the concentration of the corresponding protein denoted by $x_i$ and a list of those transitions connected to it via an arc denoted by $J_i$. To differentiate the direction of the arcs, we use positive integers to denote incoming arcs and use negative integers to denote outgoing arcs. For each transition, it maintains a rate equation $f_i(x_1, x_2, \ldots, x_3)$ which has two types described in section 2.2.1: mass-action law and Michaelis-Menten equation. Under this encoding scheme, the procedure of HFPN execution within one discrete time step (i.e. without GRN regulation) is similar to the procedure of solving a nonlinear Ordinary Differential Equation (ODE) system:

$$\frac{dx_i}{dt} = \sum_{j \in J} c_{ij} v_j$$
where $v_j = f_i(x_1, x_2, \ldots, x_3)$ is the reaction rate for which $x_i$ is either a reactant or product of, and $c_{ij}$ is its stoichiometric coefficient. To solve this ODE system, we use Numerical Integration technique chose the time interval as $\Delta t = 0.0001 \text{min}$. The HFPN execution within one discrete time step is shown in line 2 - 15 of Algorithm 2, and after that system will call regulation() to achieve the interaction between pathways and GRN.

**Algorithm 2** Simulation algorithm

```plaintext
1: for discrete time step $i = 0$ to $\infty$ do
2:   for time $t = 0$ to 240000 do
3:     for all $v_i$ do
4:       $v_i \leftarrow f_i(x_1, x_2, \ldots, x_n)$
5:     end for
6:     for all $x_i$ do
7:       for all $j \in J_i$ do
8:         if $j > 0$ then
9:           $x_i \leftarrow x_i + v_i \Delta t$
10:          else
11:            $x_i \leftarrow x_i - v_i \Delta t$
12:       end if
13:     end for
14:   end for
15: end for
16: regulation()
```

During loading the GRN portion is parsed into a Bayesian network $G = (V, E)$. For each node $g_i \in V$, it maintains its children list $C_i$ and its parent list $P_i$. In addition, each element $c_j \in C_i$ also coupled with a link strength $\eta_{ij}$. Thus we can get the CPT of each gene $g_i$ based on the formulas introduced in section 2.2.2. Algorithm 3 illustrate this procedure. Note that the $F(e_{r,c}, x)$ and $\text{Prob}(g'_i = r | g_i = c)$ refer to equation X and equation X.

**Algorithm 3** Compute Conditional Probability Table

```plaintext
1: for all gene $g_i \in V$ do
2:   $T \leftarrow$ new $|P_i|^3 \times 3$ table
3: for all entry $e_{r,c} \in T$ do
4:   for all $x$ such that $c = \lfloor f \rfloor \lceil f \rceil$, where $f \leftarrow F(e_{r,c}, x)$ do
5:     $w \leftarrow 1 - |f - c|$
6:     $p \leftarrow wp \text{Prob}(g'_i = r | g_i = c)$
7:     $s \leftarrow p + s$
8:   end for
9:   $e_{r,c} \leftarrow s$
10: end for
11: end for
```
To handle the interactions between pathways and GRN, we have two lists: direct target gene list $TL$ and feedback gene-protein pairs $FP$. When we regulate GRN, we first update corresponding link strength $\{\eta_{x\rightarrow y} \in L | y \in TL\}$ and recompute CPT. Then we topologically sort the BN and compute the probability of expression for each gene. After that, we compute the expectation of each gene appears in pair $(g_i, p_i) \in FP$, and then feedback to pathway by updating the concentration of $p_i$ (Algorithm 4).

### Algorithm 4 Regulation

1. for all $\eta_{x\rightarrow y}$ such that $y \in TL$ do
2. $\eta_{x\rightarrow y} \leftarrow k_{x,y}$
3. end for
4. compute all CPT
5. $l \leftarrow TopologySort(BN)$
6. for all $g_i \in l$ do
7. compute all $Prob(g_i)$
8. for all $g_i$ such that $(g_i, p_i) \in FP$ do
9. $p_i.x_i \leftarrow p_i.x_i \sum_{z=0}^{2} zProb(g_i = z)$
10. end for
11. end for

### 5.2 Simulation results and analysis

Before showing the simulation results, we first talk about the parameters used in Wnt model. For the pathway portion, the parameters can be divided into two classes: the initial value of protein concentrations and rate equation constant, which are summarized in Table A.2 and Table A.1 (See Appendix) correspondingly. In section 3.2.4, we have introduced the robust method for estimating parameters developed by our group; however, due to the lack of protein concentration time serials data, instead of running that estimating algorithm, the parameters for Wnt model are from literature and manual estimation. For the GRN portion, since we adopt physical approach to build GRN, the parameters (i.e. link strength) are also setting manually based on the knowledge acquired from biological publications. Furthermore, for the parameters of interaction portion, we do the similar things like our estimating algorithm which can be described as a simulation - validation - adjustment loop. Those parameters are shown in Appendix (Table A.3). Although there are artificial parameters in the model; based on the simulation results, we can provide several kinds of quantitatively analyses which may not be so precise. Here, to illustrate the features of the Hybrid modeling framework, we suggest the
5.2.1 Protein concentration time courses

The underlying assumption of our hybrid time scale is that most of the components in the systems will reach steady state very quickly, i.e. in terms of minutes. On the other hand, 4 hours should be enough for target genes to complete transcription and translation. Thus we first consider the case within one discrete time step, i.e. without the effects of GRN. The simulated protein concentration time courses are partly shown in Figure 5.1. Most of the components from both canonical and non-canonical Wnt pathway such as Axin, G-protein, etc can reach steady state, which results in the local stabilization of the systems. It shows that our assumption is reasonable, thus we can consider the interaction with of GRN next.

![protein concentration time courses](image)

There are also a few components such as β-catenin, TCF.LEF require more than 4 hours, which is conflict with our assumption (Figure 5.2). To solve this problem, one way is to extend
the time within one discrete step, but in that case, the profile of gene expression levels will be delayed comparing with the experiment data. In the living cell, it is possible that the GRN affect the pathway which is not in steady state. Thus our approach is still realistic in some sense.

As we mentioned in section 4.1.1, without the presence of Wnt signal, the concentration level of β-catenin will kept low, and the stimulation of Wnt can accumulate β-catenin in both cytoplasm and nucleus. As shown in Figure 5.2, the simulated β-catenin concentration is around 2 nM without Wnt. In contrast, in the presence of Wnt, the concentration became around 80 nM. This result is consistent with the experiment data which claim that Wnt-3A protein elevates the levels of β-catenin 4-8 fold (Willert, Epping, Pollack, Brown, & Nusse, 2002).

Next let’s consider the interactions between pathways and the GRN.

### 5.2.2 Gene expression levels

The two states of Wnt model imply that the ratio of target gene expression levels with and without Wnt stimulation is an important indicator of the effects of Wnt signal on GRN. With the probabilities of normalized gene expression levels, we can compute the expected value for each gene expression, and we believe that the expectation is proportional to the expression levels. Thus ratio of expected values will reflect the ratio of expression levels. Part A of Figure 5.4 shows the expectation of MYC gene expression level for each discrete time step. Without Wnt signal, the value is kepted low, as TCF will bind to TLE to inhibit target gene expression. After inducing Wnt, β-catenin convert that transcription repressor to transcription activator,
which up-regulate the target genes. The ratios of each time step for the two states are calculated and shown in the Part B. In a real case, for example Human EC cell, the expression fold is around 2 with and without Wnt (Willert et al., 2002), which is consistent with this simulation results.

![Figure 5.3: The comparison of β-catenin concentration](image)

![Figure 5.4: MYC gene expression](image)
5.2.3 The effects of feedback

After target genes are triggered on or off, the concentration of mRNA will be influenced which will finally affect the concentration of corresponding proteins. In Figure 5.5, the plots show the protein concentration profile of 40 hours. The decrease of receptor Fz and the increase of inhibitor Dkk can reduce the Wnt signal strength. In addition, Axin, a component of β-catenin destruction complex, has been produced in each step. Both the above two events will increase the concentration level of Axin.APC.GSK3 complex. As a result, β-catenin begin to decrease, meaning Wnt signal is fading away by its own regulation. Although the increase of TCF/LEF can slightly increase the concentration of transcription factor TCF/LEF.β-catenin, the trend of TCF/LEF.β-catenin is dropping shown in panel F of Figure 5.5.

![Graphs of Fz, Axin, Axin.APC.GSK, β-catenin, TCF/LEF, and TCF/LEF.β-catenin](image)

Figure 5.5: Protein concentration time courses with the effects of feedback

Furthermore, we can also analyze the effects of particular protein (e.g. Axin) which play an important role in the pathway, or to test hypothesis such as the mechanism of Dsh activation. Due to the constraint of the total length of the report, those content haven’t been presented.
Chapter 6

Conclusions

6.1 Summary

In this report, we have presented our work on a Hybrid modeling framework for biological networks. The main goal of this framework is to link a Signaling pathway and a Gene Regulatory Network together. We have described the details of our modeling methodology. To model the kinetics of signaling pathways, we have described a systematical way of constructing a Hybrid Functional Petri Net model. To model the interactions among genes, we have introduced a reverse-engineering approach for learning a Bayesian Network model from gene expression data. To complete the model, we also proposed a quantitative way of dealing with the interactions between pathways and a gene network, and how to estimate parameters efficiently. Then we provided a demonstration of modeling Wnt signaling pathways and its GRN. After presenting the implementation of the simulation engine, we suggested some analyses base on the simulation results.

6.2 Limitations and Future Work

This is the first work integrating a signaling pathway and gene regulatory network. We simply adopted a naive approach to model their interaction. In the living cell, not all the pathways can reach steady state in minutes, and the timing of feedback from a GRN may have a huge impact on the system dynamics. This may conflict with the underlying assumption of the hybrid time scale. Thus for future works, we should invent a mathematical way to model transcription delay
The Wnt model it is still incomplete. Firstly, there are some unknown mechanisms such as
the activation of Dsh, the binding of LRP and Axin, etc., and those debates among biologists
are still going on. In future, we hope we can have a better understanding of about those issues.
Secondly, other non-canonical Wnt pathways such as the PCP pathway can also interact with
the two pathways we modeled. To understand the roles of Wnt signaling in development and
disease more completely, we should involved other pathways to the model. The ultimate goal
about this is to implement a unified model which includes all the important pathways of a cell.
Thirdly, currently many parameters used are estimated manually which is both imprecise and
potentially biased, and due to lack of experimental data, we haven’t conducted any validation
on the parameter estimation process. After collecting enough gene expression data and protein
congcentration time series data, we will run the estimating algorithms, validate the model and
provide more analyses.
References


Reactions and Parameters

Table A.1: Rate equations with parameters used in Wnt model. The first order and second order rate constants are given in s$^{-1}$ and nM$^{-1}$s$^{-1}$ respectively while Michaelis Constants $V_{\text{max}}$ and $K_M$ are given in nM.s$^{-1}$ and nM respectively.

<table>
<thead>
<tr>
<th>No.</th>
<th>Rate Equation</th>
<th>Parameter</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$k_1$[Wnt1][Fz1.LPR5/6]</td>
<td>$k_1 = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>2</td>
<td>$k_2$[Wnt.Fz1.LPR5/6]</td>
<td>$k_2 = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>3</td>
<td>$k_3$[Wnt.Fz1.LPR5/6][Dsh]/(k_4 + [Dsh])</td>
<td>$k_3 = 0.182, k_4 = 10.0$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>4</td>
<td>$k_5$[Dsh*]</td>
<td>$k_5 = 0.0182$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>5</td>
<td>$k_7$</td>
<td>$k_7 = 0.42$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>6</td>
<td>$k_6$[Axin][APC]/(k_7 + [APC])</td>
<td>$k_6 = 0.17, k_7 = 50.0$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>7</td>
<td>$k_8$[Axin][APC]</td>
<td>$k_8 = 0.5$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>8</td>
<td>$k_9$[Axin][APC]</td>
<td>$k_9 = 0.01$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>9</td>
<td>$k_{10}$[Dsh*][Axin][APC][GSK3]</td>
<td>$k_{10} = 0.05$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>10</td>
<td>$k_{11}$[Axin][APC][GSK3]</td>
<td>$k_{11} = 0.091$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>11</td>
<td>$k_{12}$[Axin][APC][GSK3]</td>
<td>$k_{12} = 0.91$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>12</td>
<td>$k_{13}$[Axin][APCp][GSK3]</td>
<td>$k_{13} = 0.133$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>13</td>
<td>$k_{14}$[Axin][APCp][GSK3]</td>
<td>$k_{14} = 0.267$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>14</td>
<td>$k_{15}$[Axin][APCp][GSK3][catenin]</td>
<td>$k_{15} = 720.0$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>15</td>
<td>$k_{16}$[catenin][APC]</td>
<td>$k_{16} = 120.0$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>16</td>
<td>$k_{17}$[APC][catenin]</td>
<td>$k_{17} = 0.1$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>17</td>
<td>$k_{28}$[catenin]</td>
<td>$k_{28} = 0.2$</td>
<td>Estimate</td>
</tr>
<tr>
<td>18</td>
<td>$k_{29}$[catenin][APC]</td>
<td>$k_{29} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>19</td>
<td>$k_{30}$[Wnt1][Fz2]</td>
<td>$k_{30} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>20</td>
<td>$k_{31}$[Wnt5a,Fz2]</td>
<td>$k_{31} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>21</td>
<td>$k_{18}$[Axin][APCp][GSK3][catenin]</td>
<td>$k_{18} = 206.0$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>22</td>
<td>$k_{29}$[catenin]</td>
<td>$k_{19} = 0.42$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>23</td>
<td>$k_{20}$[Axin][APCp][GSK3][catenin]</td>
<td>$k_{20} = 206.0$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>24</td>
<td>$k_{21}$[catenin][Axin][APCp][GSK3]</td>
<td>$k_{21} = 6.0$</td>
<td>(Lee et al., 2003)</td>
</tr>
</tbody>
</table>

Continued on Next Page...
<table>
<thead>
<tr>
<th>No.</th>
<th>Rate Equation</th>
<th>Parameter</th>
<th>Ref.</th>
</tr>
</thead>
<tbody>
<tr>
<td>25</td>
<td>$k_{22}$[DKK][LPR5/6]</td>
<td>$k_{22} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>26</td>
<td>$k_{23}$[DKK.LPR5/6]</td>
<td>$k_{23} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>27</td>
<td>$k_{24}$[TF]/[LEF1]</td>
<td>$k_{24} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>28</td>
<td>$k_{25}$[TF].LEF1</td>
<td>$k_{25} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>29</td>
<td>$k_{26}$[TF].LEF1[nbcatenin]</td>
<td>$k_{26} = 0.1$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>30</td>
<td>$k_{27}$[TF].LEF1[nbcatenin]</td>
<td>$k_{27} = 3.0$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>31</td>
<td>$k_{32}$[Gb][Gr]</td>
<td>$k_{32} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>32</td>
<td>$k_{33}$[Gb]</td>
<td>$k_{33} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>33</td>
<td>$k_{34}$[Gb][Ga]</td>
<td>$k_{34} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>34</td>
<td>$k_{35}$[Wnt5a.Fz2][Gabr]/(k_{36} + [Gabr])</td>
<td>$k_{35} = 0.2$, $k_{36} = 50.0$</td>
<td>Estimate</td>
</tr>
<tr>
<td>35</td>
<td>$k_{37}$[PCL][Gb]/(k_{38} + [Gb])</td>
<td>$k_{37} = 0.2$, $k_{38} = 50.0$</td>
<td>Estimate</td>
</tr>
<tr>
<td>36</td>
<td>$k_{39}$[PCL*]</td>
<td>$k_{39} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>37</td>
<td>$k_{40}$[PCL*][PIP2]</td>
<td>$k_{40} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>38</td>
<td>$k_{41}$[DAG][IP3]</td>
<td>$k_{41} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>39</td>
<td>$k_{42}$[DAG][PKC]</td>
<td>$k_{42} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>40</td>
<td>$k_{43}$[PKC*]</td>
<td>$k_{43} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>41</td>
<td>$k_{44}$[IP3]</td>
<td>$k_{44} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>42</td>
<td>$k_{45}$[IP2]</td>
<td>$k_{45} = 0.05$</td>
<td>Estimate</td>
</tr>
<tr>
<td>43</td>
<td>$k_{46}$[IP3]</td>
<td>$k_{46} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>44</td>
<td>$k_{47}$[IP4]</td>
<td>$k_{47} = 0.05$</td>
<td>Estimate</td>
</tr>
<tr>
<td>45</td>
<td>$k_{48}$[IP3][Ca2+ER]</td>
<td>$k_{48} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>46</td>
<td>$k_{49}$[Ca2+]</td>
<td>$k_{49} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>47</td>
<td>$k_{50}$[Ca2+][CamKII]</td>
<td>$k_{50} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>48</td>
<td>$k_{51}$[CamLII*]</td>
<td>$k_{51} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>49</td>
<td>$k_{52}$[Ca2+][Calcineurin]</td>
<td>$k_{52} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>50</td>
<td>$k_{53}$[Calcineurin*]</td>
<td>$k_{53} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>51</td>
<td>$k_{54}$[Calcineurin*][NFATp]</td>
<td>$k_{54} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>52</td>
<td>$k_{55}$[NFAT]</td>
<td>$k_{55} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>53</td>
<td>$k_{56}$[NFAT]</td>
<td>$k_{56} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>54</td>
<td>$k_{57}$[nNFAT][GSK]</td>
<td>$k_{57} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>55</td>
<td>$k_{58}$[nNFATp]</td>
<td>$k_{58} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>56</td>
<td>$k_{59}$[nNFATp]</td>
<td>$k_{59} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>57</td>
<td>$k_{60}$[nNFAT]</td>
<td>$k_{60} = 0.0010$</td>
<td>Estimate</td>
</tr>
<tr>
<td>60</td>
<td>$k_{67}$[Dshp]</td>
<td>$k_{67} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>62</td>
<td>$k_{66}$[PDK*][Dsh]/(k_{69} + [Dsh])</td>
<td>$k_{69} = 0.1$, $k_{69} = 10.0$</td>
<td>Estimate</td>
</tr>
<tr>
<td>66</td>
<td>$k_{109}$[CamLII*][LEF1]/(k_{108} + [LEF1])</td>
<td>$k_{109} = 0.1$, $k_{108} = 20.0$</td>
<td>Estimate</td>
</tr>
<tr>
<td>67</td>
<td>$k_{71}$[LEF1p]</td>
<td>$k_{71} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>68</td>
<td>$k_{73}$[bcatenin]</td>
<td>$k_{73} = 2.57E - 4$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>77</td>
<td>$k_{107}$</td>
<td>$k_{107} = 8.2E - 5$</td>
<td>(Lee et al., 2003)</td>
</tr>
<tr>
<td>80</td>
<td>$k_{85}$[Ga][PDE]/(k_{86} + [PDE])</td>
<td>$k_{85} = 1.0$, $k_{86} = 50.0$</td>
<td>Estimate</td>
</tr>
<tr>
<td>81</td>
<td>$k_{87}$[PDE*]</td>
<td>$k_{87} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>82</td>
<td>$k_{88}$[PDE*][cGMP]/(k_{89} + [cGMP])</td>
<td>$k_{88} = 1.0$, $k_{89} = 50.0$</td>
<td>Estimate</td>
</tr>
<tr>
<td>83</td>
<td>$k_{90}$[cGMP+ER]</td>
<td>$k_{90} = 0.01$</td>
<td>Estimate</td>
</tr>
<tr>
<td>92</td>
<td>$k_{90}$[Fz1][LPR5/6]</td>
<td>$k_{90} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>93</td>
<td>$k_{100}$[Fz1.LPR5/6]</td>
<td>$k_{100} = 0.1$</td>
<td>Estimate</td>
</tr>
<tr>
<td>Reactant</td>
<td>Concentration (nM)</td>
<td>Ref</td>
<td></td>
</tr>
<tr>
<td>------------</td>
<td>-------------------</td>
<td>-------------------------------</td>
<td></td>
</tr>
<tr>
<td>Wnt1</td>
<td>10.0</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>Fz1</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>Dsh</td>
<td>100.0</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>Axin</td>
<td>0.02</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>APC</td>
<td>100.0</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>GSK3</td>
<td>50.0</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>bcatenin</td>
<td>34.0</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>TCF</td>
<td>15.0</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>bcateninp</td>
<td>1.0</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>DKK</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>LPR5/6</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>LEF1</td>
<td>15.0</td>
<td>(Lee et al., 2003)</td>
<td></td>
</tr>
<tr>
<td>Fz2</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>G-protein</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>PCL</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>PIP2</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>PKC</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>Ca2+ER</td>
<td>100.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>CamKII</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>Calcineurin</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>NFAT</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>GSK</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>Naked</td>
<td>5.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>PDE</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>cGMP</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
<tr>
<td>Wnt5a</td>
<td>10.0</td>
<td>Estimate</td>
<td></td>
</tr>
</tbody>
</table>

Table A.2: Initial concentration of the pathway components
<table>
<thead>
<tr>
<th>From</th>
<th>To</th>
<th>Parameter</th>
<th>Type</th>
</tr>
</thead>
<tbody>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>MYC</td>
<td>0.23</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>JUN</td>
<td>0.25</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>CCND1</td>
<td>0.18</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>PPARD</td>
<td>0.21</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>MM7</td>
<td>0.20</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>VEGF</td>
<td>0.17</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>FOSL1</td>
<td>0.17</td>
<td>And</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>TCF1</td>
<td>0.12</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>LEF1</td>
<td>0.09</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>Axin2</td>
<td>0.19</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>DKK</td>
<td>0.20</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>BTRC</td>
<td>0.15</td>
<td>Or</td>
</tr>
<tr>
<td>TCF/LEF.bcatanin</td>
<td>ID</td>
<td>0.09</td>
<td>Or</td>
</tr>
<tr>
<td>NFAT</td>
<td>COX2</td>
<td>0.34</td>
<td>Or</td>
</tr>
<tr>
<td>NFAT</td>
<td>IL-13</td>
<td>0.32</td>
<td>Or</td>
</tr>
<tr>
<td>NFAT</td>
<td>FASL</td>
<td>0.32</td>
<td>Or</td>
</tr>
<tr>
<td>NFAT</td>
<td>IL-2</td>
<td>0.34</td>
<td>And</td>
</tr>
<tr>
<td>NFAT</td>
<td>GM-SSF</td>
<td>0.28</td>
<td>And</td>
</tr>
<tr>
<td>TCF1</td>
<td>Tcf</td>
<td>0.91</td>
<td>-</td>
</tr>
<tr>
<td>AXIN2</td>
<td>Axin</td>
<td>0.97</td>
<td>-</td>
</tr>
<tr>
<td>DKK</td>
<td>Dkk</td>
<td>0.88</td>
<td>-</td>
</tr>
<tr>
<td>BTRC</td>
<td>β-TrCP</td>
<td>0.73</td>
<td>-</td>
</tr>
<tr>
<td>FZ</td>
<td>Fz1</td>
<td>0.92</td>
<td>-</td>
</tr>
<tr>
<td>FZ2</td>
<td>Fz2</td>
<td>0.65</td>
<td>-</td>
</tr>
</tbody>
</table>

Table A.3: Interactions between pathway and GRN