2008 American Control Conference
Westin Seattle Hotel, Seattle, Washington, USA
June 11-13, 2008

Model Reduction and Consistency in Discrete Event Representation of Biological Systems

Shauheen Zahirazami, Shahin Hashtrudi-Zad and Amir G. Aghdam

Abstract— **The recent developments in biological experiments have awarded the research community with valuable information, which describe finely regulated systems that govern the cell dynamics. One of the greatest challenges, however, remains to represent this extensive amount of knowledge in a proper way that can be used in simulations, and validated automatically, in order to ultimately achieve a desired behavior for the system (cell) under control. Many tools and techniques have been proposed in the literature to address this important problem. In this paper, the use of Petri nets for knowledge representation is investigated. Two algorithms are provided to construct Petri net models for cell dynamics using data available in public domain biological database. The first algorithm generates a low-level model capturing protein-protein interactions and the second, produces a high-level model which describes pathway sequences and is considerably easier to analyze. A concept of consistency between the two models is introduced and a test for consistency is provided.**

I. Introduction

From a control engineering point of view, a cell of an organism can be considered as a dynamical system, for which a proper controller can be designed in such a way that it ensures desirable behavior. However the first step to design a proper controller would be to model the system with an acceptable accuracy. A very useful survey in applications of control in molecular systems biology is presented in [1] that illustrates the basic biological concepts and describes some open questions in dynamics and control of this type of systems. One of the building blocks of biological networks is protein interaction. Various types of interactions are orchestrated to achieve different objectives. Certain subsets of these interactions which construct a subsystem of the metabolic network are known as pathways. Due to the large size of the model of the cell, it is obvious that a systematic way of representing the knowledge of the model is highly desirable. Many attempts have been made to capture the dynamics of such systems by using differential equations [2], [3]. While these models have proven to be very useful, the complexity of the nonlinear model obtained grows rapidly by the number of protein interactions involved. A detailed approach which takes the entire cell dynamics into consideration will require a huge model which is impossible to analyze. Furthermore constructing such a model requires fine knowledge of many coefficients that are either unavailable or very hard to obtain. One approach to simplify the model with such a scale is to omit the continuous behaviors, and construct a discrete event model [4], [5], [6] or perhaps, to use a hybrid automaton approach [7], [8].

Developing early detection methods, accurate prognosis and effective therapies for numerous diseases such as cancer requires a system level understanding of the cell. In this work, an algorithm is proposed to convert the information about protein interactions in pathway block diagrams, available in public domain biology knowledge bases, into Petri-nets (one particular type of discrete-event systems). This model which describes protein-protein interactions will be referred to as the low-level model. Such discrete-event models can be extensively used in new drug discovery, along with computer simulation results, to complement or replace actual experiments. Note that these experiments are very expensive in general, can raise ethical issues, and may take a very long time to provide results.

Although the obtained model is easier to analyze compared to a continuous model, it is still very complicated at the level of the whole cell. The second contribution of this paper is to propose a new high-level model to represent the sequence of pathway activations, rather than individual protein-protein interactions. An algorithm is presented to construct the high-level model from the pathway relations obtained from public domain knowledge bases.

Next the consistency of the low-level and high-level models is examined. Loosely speaking, the high-level model is consistent with the low-level, if every pathway sequence in the low-level is also possible in the high-level model. A necessary and sufficient condition for modeling consistency is provided. The abstraction in the high-level model is meant to simplify the understanding of the dynamics of the system while maintaining the important characteristics of the low-level fine model. The high-level model may be used in the development of control systems to affect the behavior of the system. Note that the two high-level and low-level models obtained this way have different sets of events and obviously different languages, and a direct correspondence of the places and transitions are not necessarily available.

The rest of this paper is organized as follows. In Section II, some preliminary introduction is provided. In Section III, a systematic method is introduced to construct the Petri net model from a pathway diagram. Section IV offers an algorithm to create a high-level abstract model for the system from pathway-to-pathway relation diagrams. The consistency of abstract and original models are investigated in Section V. Finally, some concluding remarks are drawn in Section VI.

The authors are with the Department of Electrical and Computer Engineering, Concordia University, Montréal, Québec, Canada H3G 1M8 s_zahira, shz, aghdam@ece.concordia.ca

This work has been supported by Natural Sciences and Engineering Research Council of Canada (NSERC) under grants RGPIN-227688-06 and RGPIN-262127-07.

II. Preliminaries

A. Some Biology Background

While in this article discussing biological details has been avoided, yet it is necessary for the reader to have a basic understanding of some aspects of cell dynamics.

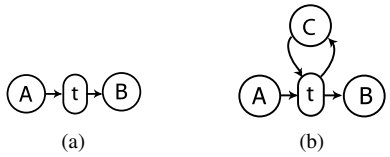
The basic building blocks of the dynamics of the cell are the individual protein to protein interactions. Almost every living organism has a DNA that serves as a genetic blueprint for construction of various proteins. Biologists and bioinformaticians are trying to discover different relationships between more than half a million proteins, as one of the most interesting and important areas of research in the post-genomic era. It is well-known that the state of the cell evolves when certain proteins interact with each other and hence form different products, which turn out to be proteins too, and might involve in some other interactions.

Combination of consequent interactions builds a *metabolic network*. While the exact definition of a *pathway* inside a metabolic network is not straightforward, it basically consists of a subset of the network that has a distinguished function. A cell constantly receives and responds to chemical signals from its environment. A signal is initiated when some extracellular signaling molecule (ligand) binds with its respective cell surface receptor. Signaling molecules inside the cell then interact to transduce the signal into cellular responses. Specific collections of interconnected interactions in a network are often referred to as pathways.

B. Petri nets

In a Petri net model the edges are divided into two groups, namely places and transitions. No two places are connected directly to each other; no two transitions are connected to each other either. A transition can "fire" only if the places connected to it have enough tokens to overcome the weight of the arcs going from those places to the transitions. When a transition fires, it takes some tokens from its inbound places and puts the tokens into the outbound places.

Let $\mathbb{Z}^+$ denote the set of non negative integers. A labelled Petri net is a six tuple $G = (P,T,F,L,M_0,\Omega)$ where P and T are finite sets of places and transitions, respectively. F is a flow function defined as $F : (P \times T) \cup (T \times P) \to \mathbb{Z}^+$ (for $F(x,y) > 0$, there is an arc from x to y denoted by $A(x,y)$ with *multiplicity* $F(x,y)$), and $L : T \to \Omega$ is a labeling, (labeling attaches an action name to each transition). The definition of L can also be extended to $L : T^* \to \Omega^*$, where Ω^* denotes the set of finite sequences of the elements of Ω, including the zero length string ε. For the Petri nets in this paper, $T = \Omega$ and L is the identity map. M_0 is an initial marking of the Petri net, where a marking M is a function that gives the number of tokens in each place defined as $M : P \to \mathbb{Z}^+$. A transition t is enabled at a marking M, if for all $p \in P$, $M(p) \geq F(p,t)$ and is denoted by $M \xrightarrow{t}_G$. An enabled transition like t may fire at a marking M yielding the marking M' that is denoted by $M \xrightarrow{t}_G M'$, where $M'(p) = M(p) - F(p,t) + F(t,p)$ for all $p \in P$. Also for any $a \in \Omega$, $M \xrightarrow{a}_G$ means $M \xrightarrow{t}_G$ for some t with $L(t) = a$. The reachability set of a Petri net G is defined as $\Re(G) = \{M | \exists \sigma \in T^*, M_0 \xrightarrow{\sigma}_G M\}$. A place $p \in P$ is unbounded if for every $k \in \mathbb{Z}^+$, there exists $M \in \Re(G)$ such that $M(p) > k$. The language generated by the Petri net G is defined as $\ell(G) = \{\omega \in \Omega^* | M_0 \xrightarrow{\omega}_G\}$.



(a) (b)

Fig. 1: (a) Translocation $A \to B$, (b) Simplified enzymatic activation (inhibition) $A \xrightarrow{C} B$

C. Modeling protein interactions

Assume that protein A and B can interact and produce the compound AB, there would be two directed arcs from protein A and protein B to interaction T and one arc from interaction T to the product protein AB. As an example, Figure 1b shows an enzymatic activation, where the presence of enzyme C in the environment (i.e. tokens in place C) is facilitating the chemical process from A to B, and C remains intact after the process. While modeling protein interactions may seem straight forward, in some cases, the definition of the pathway could be fuzzy, and certain challenges have to be faced in order to use publicly available biological knowledge [9]; however in this paper it is assumed that an exact set of interactions are obtained from public domain biological literature such as Kyoto Encyclopedia of Genes and Genomes (KEGG). KEGG is a bioinformatics resource, developed as part of the research projects in the Kanehisa Laboratory of Kyoto University Bioinformatics Center. In order to capture maximum knowledge possible, the pathways are represented by a block diagram with different types of connections. In order to understand different types of interactions and events which are represented using various forms of arcs, a legend table has to be used. This diagram specifies the starting points of the pathway (i.e., cell surface receptors) through other interactions and up to where the pathway completes, and triggers other pathways. A *pathway-to-pathway relationship diagram* showing the causal effect of pathways on each other can also be obtained in terms of tables of consequent pathways as presented in Section IV.

In this work, we investigate a systematic method to obtain a Petri net model from public domain knowledge. A model reduction algorithm will thence presented to tackle the large scale nature of the model. A supervisor can then be developed for the reduced-order model as a basis for possible drug design or therapies. The designed supervisor controller could essentially be an autonomous molecular computer [10], that would be used for anti-viral therapies, where systematic gene delivery to target and kill metastases is applied using genetically modified viruses or stem cells [11].

III. Development

As discussed in the previous section, various pathway diagrams showing protein-protein interactions have been obtained and are available from a number of databases such as KEGG. Figure 2 shows the Apoptosis pathway obtained from

KEGG. Note that although the sketch has been manually curated and hence, is highly accurate and readable to human, the representation is not flexible enough to be taken as a machine knowledge representation tool.

In this section, an algorithm is proposed that can translate a pathway block diagram such as Figure 2, to a Petri net model that efficiently describes the pathway knowledge. The resulting Petri net can be understood and updated automatically, and queried by a program. Furthermore, it can be studied using the tools and methods readily available for the analysis of Petri nets.

Assume that all protein interactions are translocations (Figure 1a). Other types of reactions can be easily accommodated as discussed later in Remark 1.

Algorithm 1:

Step 1: Construct a set $\mathbf{I}$, consisting of the initial starting proteins in the pathway.

Step 2: Set $j = 1$.

Step 3: Choose a p_j in $\mathbf{I}$, and set $\mathbf{I} = \mathbf{I} - \{p_j\}$.

Step 4: Add a place p_j in the Petri net set of places P if it has not already been added to this set.

Step 5: Construct $\mathbf{L}$, the set of all proteins directly connected to p_j with an incoming arc in the pathway.

Step 6: Set $\mathbf{I} = \mathbf{I} \cup (\mathbf{L} - P)$.

Step 7: For all p_k in L if $p_k \notin P$, add a place p_k in P.

Step 8: For all p_k in L, create transition $t_{j,k}$ and corresponding arcs $A(p_j, t_{j,k})$ and $A(t_{j,k}, p_k)$.

Step 9: If $\mathbf{I} \neq \emptyset$, then set $j = j + 1$ and go to *Step 3*; else stop the algorithm. ■

The output of the algorithm is a low-level model G_{lo}. In the resulting Petri net, the tokens in each place would correspond to the concentration of a protein and the transitions would represent protein-to-protein interactions.

Remark 1: Step 8 of *Algorithm 1* can easily be rewritten to: (i) incorporate a case decision maker that constructs the transitions based on a lookup table, for different types of interactions (i.e. different arrows in Figure 2); and (ii) set different arc weights based on rate of participation of different proteins in each interaction. ■

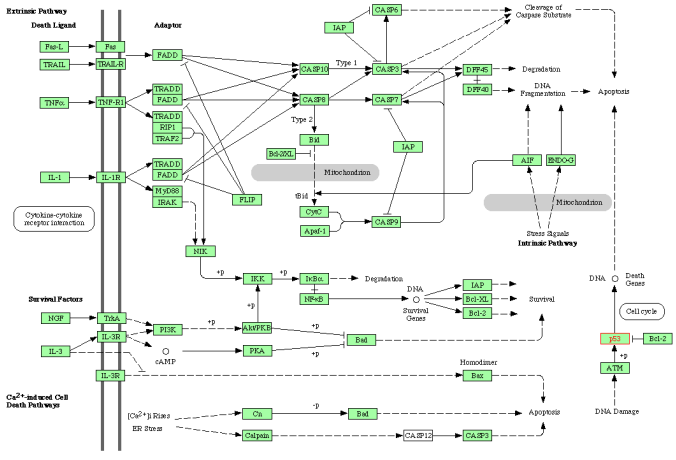


Fig. 2: Apoptosis pathway obtained from the KEGG

Example 1: Using *Algorithm 1* a simple artificial protein interaction network consisting of ten proteins can be processed to obtain the Petri net in Figure 3b.

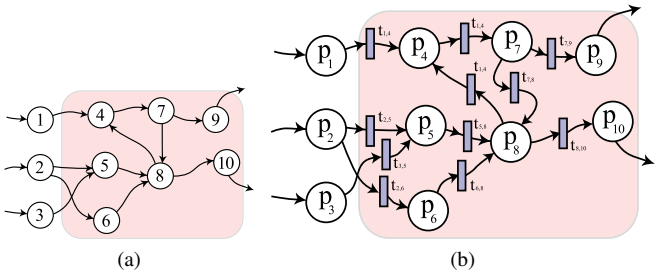


Fig. 3: Input and resulting Petri net in Example 1

Following *Algorithm 1*, Petri net models can be constructed for various pathways, which in turn, can be combined to form a large model. This model can be used in the analysis of cell behavior or possibly be used in the development of supervisors to enforce a desired behavior (e.g., activation of the desirable pathways and disablement of the undesirable ones). Even though this model is extremely useful, the number of places can get very large, as we add different proteins to the model. Therefore, it is desirable to construct a more abstract compact model that, for instance, describes cell dynamics at the pathway level. A pathway level of abstraction is particularly useful since it is not uncommon for the desired cell behavior (control specification) to be expressed in terms of pathways, rather than the individual proteins.

IV. Reduced Model

In order to proceed with the abstraction, it is desired to identify the subsystems within the obtained model. The approach presented here considers each pathway as a subsystem with a different functionality. However, instead of directly reducing the previously obtained low-level model, the abstract (high-level) model is derived using the *pathway-to-pathway relation diagram* available in databases such as KEGG. For instance, in KEGG the list of all pathways triggered by each pathway is provided. In the case of the Apoptosis pathway, for example, the list includes, Cell communication, Calcium signaling pathway, Cell cycle, Colorectal cancer, Pancreatic cancer.

Essentially, the relation among pathways is represented in the form of a digraph which may very well have loops. Furthermore, the causal effects of the pathways may change the dynamics of the whole system. One advantage of not developing a high-level model from the low-level one is that the consistency of the two models can be examined through a proper comparison method. (Details will be provided in the next section.) Inconsistencies, if any, point to deficiencies or inaccuracies in the information used for setting up the low-level and high-level models and to the possibility of future experimental work necessary to correct the inconsistencies.

Here, an algorithm is proposed for translating pathway-to-pathway diagrams into Petri net models. In the resulting Petri net, places represent pathways and transitions represent completion of one pathway and preparation of the other.

Hence, a pathway is active when there are tokens in the respective place. Notice that there is no specific starting point in the pathway-to-pathway relation diagram, and the algorithm finishes only when all the knowledge has been captured.

Algorithm 2:

Step 1: Set $j = 1$ and $\mathbf{I} = \{p_1\}$.

Step 2: Choose a random pathway p_j from $\mathbf{I}$.

Step 3: Remove the chosen p_j from $\mathbf{I}$.

Step 4: Add a place corresponding to p_j in the Petri net set of places P and a transition t_j with an arc going from p_j to t_j.

Step 5: From the pathway relation graph, construct a set $\mathbf{L}$ of all the pathways resulting from p_j (i.e., with an incoming arc from p_j) and a set B containing the pathways that result in p_j (i.e. with an arc incoming to p_j).

Step 6: Set $\mathbf{M} = \mathbf{L} \cap P$, and $\mathbf{B} = \mathbf{B} \cap P$, and $\mathbf{L} = \mathbf{L} - P$.

Step 7: If $\mathbf{B} = \emptyset$ then go to *Step 10*.

Step 8: Choose any p_k from $\mathbf{B}$, and add an arc $A(t_k, p_j)$.

Step 9: Remove p_k from $\mathbf{B}$, and then go to *Step 7*.

Step 10: If $\mathbf{M} = \emptyset$, then go to *Step 13*.

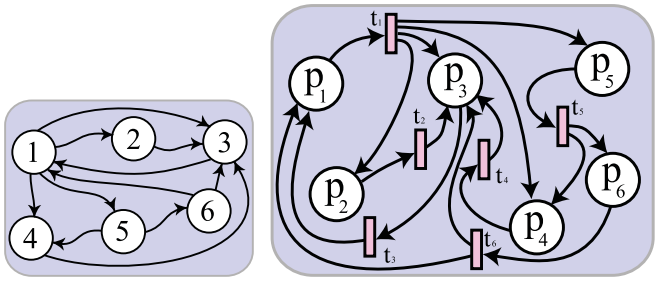
Step 11: Choose any p_k from $\mathbf{M}$, and add an arc $A(t_j, p_k)$.

Step 12: Remove p_k from $\mathbf{M}$, and then go to *Step 10*.

Step 13: Set $\mathbf{I} = \mathbf{I} \cup \mathbf{L}$.

Step 14: If $\mathbf{I}$ is not empty, then increment j and go to *Step 2*; otherwise stop the algorithm. ■

Example 2: Using *Algorithm 2* A simple artificial pathway-to-pathway relation diagram consisting of six pathways can be processed to obtain the Petri net in Figure 4b. ■



(a) (b)

Fig. 4: Input and resulting Petri net in Example 2

The reduced high-level model G_{hi} obtained from *Algorithm 2* does not consider individual transitions of the involved molecules and proteins; however, it describes the overall advancement of the pathways in the cell. Therefore, based on an analysis on the simplified high-level model, one would be able to draw valuable conclusions about the more complex low-level model.

A high-level model would specifically be useful when designing a controller. It would ultimately be possible to identify which chain of interactions should be interrupted or which pathway has to proceed. It would also be possible to examine the system using known fault recovery methods of discrete event systems [12] to come up with a set of hypotheses that would suggest what missing interactions facilitate or interrupt a certain pathway.

V. Modeling and control Consistency

Since the two models G_{hi} and G_{lo} have been obtained from two different knowledge sources, it is now desired to investigate whether the language behaviors of the two models can be directly compared or not. G_{hi} is an abstract model of the system and is supposed to capture the information about the sequences of pathway activations in the cell. Referring to the aforementioned property of modeling *consistency*, it is expected, at least, that every sequence of pathway activation that is feasible in G_{lo} be also feasible in G_{hi}. An important feature of the problem of comparing G_{lo} and G_{hi} is that they have different sets of transitions, and thus, their languages are defined over different alphabets.

To begin the discussion, the previous notation is slightly modified such that the transitions and places of G_{lo} and G_{hi} are not confused. Let P_{lo}, T_{lo} and P_{hi}, T_{hi} designate the place and transition sets of the low-level and high-level models, respectively. The places in G_{lo} (elements of P_{lo}) are shown by p and those in G_{hi} (which correspond to the pathways) are denoted by v. The low-level transitions are denoted by t and the high-level ones by l. The low-level and high-level markings are represented by M_{lo} and M_{hi}, respectively.

In order to study the consistency of the models, define the following relations.

Definition 1: Let $R_T \subseteq T_{lo} \times P_{hi}$ be a binary relation such that a pathway $v \in P_{hi}$ is in relation with an interaction $t \in T_{lo}$ (written as tR_Tv or $(t,v) \in R_T$) if and only if t can happen while pathway v is active. ■

Definition 2: Let $R_P \subseteq P_{lo} \times P_{hi}$ be a binary relation such that the pathway $v \in P_{hi}$ is in relation with $p \in P_{lo}$ (pR_Pv or $(p,v) \in R_P$) if and only if p is a protein that participates in an interaction that belongs to pathway v. ■

Denote the set of *boundary transitions* that can trigger new pathways by T_b:

$$T_b = \{t \in T_{lo} | \exists v_1, v_2 \in P_{hi}, p \in P_{lo}, v_1 \neq v_2 : (t,v_1) \in R_T, (t,v_2) \notin R_T, (p,v_2) \in R_P, F(t,p) \neq 0\}.$$

The following two assumptions are made about the boundary transitions.

Assumption 1: Each pathway can have at most one boundary transition.

Suppose Assumption 1 does not hold, and for instance, a pathway v can trigger another pathway such as v'_1 with boundary transition t_1 and pathway v'_2 with boundary transition t_2. In this case, in order to be able to trace pathway activations at the high-level model, the information that v may trigger v'_1 or v'_2 but not both at the same time, should be available in the high-level knowledge source, which may not be true in general (as in the case of pathway-to-pathway relationship diagram of KEGG). To ensure the assumption remains valid, one may refine v, by splitting it into two pathways: one triggering v'_1 and the other v'_2.

Assumption 2: A boundary transition belongs to one pathway only (no common boundary transitions).

Suppose Assumption 2 does not hold, and for instance, pathways v_1 and v_2 have a common boundary transition

t which triggers v'. In order to be able to correctly trace pathway activations at the high-level, the information concerning the combination of the two pathways v_1 and v_2 for triggering v' must be available in the high-level knowledge source which may not be true in general (as in the case of the KEGG pathway-to-pathway relationship diagram). In order to keep the assumption valid, one can combine pathways v_1 and v_2 in a larger section with one boundary transition t.

As mentioned in Section II-B, the marking of a Petri net describes the state of the Petri net (as a dynamic system). In this case, any marking of G_{lo}, such as M_{lo}, is a possible vector of protein and molecule concentrations in the cell at a specific time. The concentration levels in M_{lo} will cause different pathways to become active or inactive. For a given concentration of proteins (i.e. tokens in the places) of the pathway v, the maximum number of times that the pathway v can trigger its boundary transition in the absence of any external transitions (i.e. transitions not belonging to the pathway) depositing tokens in the pathway places, is considered in this paper as the number of times the pathway v is activated.

Specifically, let $G_{lo,v}$ denote the sub-Petri net of G_{lo} consisting of places and transitions of pathway v, and $M_{lo,v}$ denote the marking of $G_{lo,v}$. Furthermore, let $\ell(G_{lo,v}, M_{lo,v})$ denote the language generated by $G_{lo,v}$ starting from $M_{lo,v}$ and t_b (the boundary transition of $G_{lo,v}$). Then the number of times v is considered activated at $M_{lo,v}$, (or simply the *activation* of v at $M_{lo,v}$), is

$$\alpha(v, M_{lo,v}) = \max\{n \in \mathbb{N} | \exists s \in \ell(G_{lo,v}, M_{lo,v}) : s \text{ contains } n \text{ instances of } t_b\}.$$

The number $\alpha(v, M_{lo,v})$ is, in fact, a measure of the potential of the pathway v to trigger other pathways in the absence of any external transitions. Note that from the definition of α that if as a result of an internal transition $M_{lo,v}$ becomes $M'_{lo,v}$, then $\alpha(v, M_{lo,v}) \geqslant \alpha(v, M'_{lo,v})$. In other words, in the absence of external transitions, the activation potential does not increase.

At any given time, the set of the activations of all pathways simply translates to a marking for the high-level model. In other words it is possible to define a function $\theta : \mathscr{M}_{lo} \to \mathscr{M}_{hi}$ that maps any low-level marking M_{lo} to a high-level marking $M_{hi} = \theta(M_{lo})$ with $M_{hi}(v) = \alpha(v, M_{lo,v})$.

$\theta(M_{lo})$ can be regarded as what the high-level tokens should be based on the concentrations of proteins and the corresponding state (active/inactive) of the pathways.

Let M^0_{lo} and $M^0_{hi} := \theta(M^0_{lo})$ denote the initial markings of G_{lo} and G_{hi}. M^0_{lo} is the vector of initial protein and molecule concentrations (possibly at the beginning of cell's life), and M^0_{hi} gives the corresponding active pathways.

It is desired now to study the cell behavior for variety of initial states (markings). Let Φ^0_{lo} and Φ^0_{hi} denote the set of all the initial markings of interest.

In the study of the consistency between the low-level and high-level models, the low-level markings that are reached as a result of boundary transitions play an important role in establishing synchrony between the models.

Definition 3: A low-level marking M_{lo} is called a *synchronous reachable marking* with respect to the initial marking M^0_{lo} if M_{lo} is either equal to M^0_{lo}, or M_{lo} can be reached from M^0_{lo} through a sequence of transitions in which the last transition is a boundary transition. ■

Denote the set of synchronous reachable markings with $R_s(M^0_{lo})$. One can write

$$R_s(M^0_{lo}) = \{M^0_{lo}\} \cup \{M_{lo} | \exists s \in T^*, t_b \in T_b : M^0_{lo} \xrightarrow{st_b}_{G_{lo}} M_{lo}\}.$$

Notation 1: For two low-level markings M_{lo} and M'_{lo}, and a boundary transition $t_b \in T_b$, the notation $M_{lo} \xRightarrow{t_s}_{G_{lo}} M'_{lo}$ is used if either $M_{lo} \xrightarrow{t_s}_{G_{lo}} M'_{lo}$ or $M_{lo} \xrightarrow{t_1,\ldots,t_n t_s}_{G_{lo}} M'_{lo}$, where $t_i \notin T_b (1 \leqslant i \leqslant n)$. ■

Therefore, $M_{lo} \xRightarrow{t_s}_{G_{lo}} M'_{lo}$ if M'_{lo} can be reached from M_{lo} through a sequence of transitions in which the last transition is the boundary transition t_s and the rest of the transitions are not boundary transitions.

The notion of *consistency* will formally be introduced next.

Definition 4: G_{hi} is called *consistent* with G_{lo} with respect to an initial marking M^0_{lo} if for any $M_{lo}, M'_{lo} \in R_s(M^0_{lo})$, the relation $M_{lo} \xRightarrow{t_b}_{G_{lo}} M'_{lo}$ for some $t_b \in T_b$ implies that (i) there exists a reachable high-level marking M_{hi} with $M_{hi} \geqslant \theta(M_{lo})$, and (ii) for any such high-level marking, there exists another high-level marking M'_{hi}, with $M'_{hi} \geqslant \theta(M'_{lo})$ such that with a single high-level event $M_{hi} \to_{G_{hi}} M'_{hi}$. ■

Note that $M_{hi} \geqslant \theta(M_{lo})$ means every element of M_{hi} is greater than or equal to the corresponding element in $\theta(M_{lo})$.

Definition 5: If G_{hi} is consistent with G_{lo} for every initial marking in the set of initial markings Φ^0_{lo}, then G_{hi} is *consistent* with G_{lo}. ■

The models will not be consistent if a behavior (i.e. pathway activation sequence) at the low-level is not captured in the high-level model.

Remark 2: It follows from Definition 4 that in the low level model a pathway may not be able to fire a boundary transition triggering another pathway, while in the high level the corresponding pathway may have enough tokens to fire the transition to trigger the other pathway. (This may happen when $M'_{hi} > \theta(M'_{lo})$). ■

It is desired now to develop a procedure for verifying consistency. Equivalence and refinement relations based on bisimulation and simulation relations between two DES models look for the pairs of states (Petri net markings) that can be reached after the occurrence of identical strings. However, in the problem discussed here, G_{lo} and G_{hi} have completely different transition sets (i.e. alphabet). To study and verify the property of consistency, a mapping from the sequences of the low-level model to those of the high-level one is needed. This mapping is performed using an intermediate Petri net G_{mi} whose places belong to G_{hi} and transitions are from G_{lo}. Specifically, from the procedure of construction of G_{hi} (Section IV), it follows that for every high-level transition t, there exists a unique input place (pathway) v with an arc from v to t (i.e. $F_{hi}(v,t) \neq 0$).

Definition 6: G_{mi} is obtained from G_{hi} by simply replacing every transition t in G_{hi} by the boundary transition of its

input place (i.e. pathway) v. If P_{mi} and T_{mi} denote the set of places and transitions of G_{mi}, then $P_{mi} = P_{hi}$ and $T_{mi} \subseteq T_b$. Furthermore, the initial marking M^0_{mi} is defined to be M^0_{hi}, which is equal to $\theta(M^0_{lo})$. ■

It is now necessary to define the following relation between the languages generated by the two models.

Definition 7: Consider two DES G_1 and G_2, defined over alphabets Ω_1 and Ω_2 with $\Omega_1 \subseteq \Omega_2$. The language $\ell(G_1)$ summarizes (*in the sense of abstraction*) $\ell(G_2)$ (with respect to alphabet Ω_1), (which is denoted by $\ell(G_1) \precsim \ell(G_2)$) if $\Theta_{\Omega_1}(\ell(G_2)) \subseteq \ell(G_1)$, where $\Theta_{\Omega_1} : \Omega_2 \to \Omega_1$ is the natural projection of Ω_2 on Ω_1. ■

The natural projection Θ drops all characters that do not belong to Ω_1 (i.e., characters in $\Omega_2 - \Omega_1$). In definition 7, model G_1 can be regarded as an abstraction of G_2, providing summary information about the occurrence of the events of Ω_1 only. Note that this abstraction is conservative in the sense that $\ell(G_1)$ may contain sequences that are not the projection of any sequence in $\ell(G_2)$. The following Theorem establishes a necessary and sufficient condition for the consistency of G_{hi} with G_{lo}, with respect to a given initial condition M^0_{lo}.

Theorem 1: G_{hi} is consistent with G_{lo} with respect to an initial marking M^0_{lo} if and only if $\ell(G_{mi}) \precsim \ell(G_{lo})$. ■

Remark 3: If $\ell(G_{mi}) \precsim \ell(G_{lo})$ did not hold, a possible conclusion would be that, incomplete knowledge has been used in constructing the model in specific pathways. Such cases would be candidates for further lab tests.

Remark 4: The condition in Theorem 1 involves projection of languages and can be verified (assuming bounded Petri nets and thus, regular languages), by for example converting the Petri nets to automata and using operations on automata.

Example 3: To demonstrate the results, suppose that two pathways $v_1(\{p_1,p_2,p_3\},\{t_1,t_2\})$ and $v_2(\{p_4,p_5\},\{t_3\})$ simply consisting of five proteins and three interactions are defined as depicted in Figure 5. As can be observed, the transition t_3 is a boundary transition between the pathways v_1 and v_2. An initial marking $M^0_{lo} = [p_1,p_2,p_3,p_4,p_5] = [0,1,1,1,1]$ is given for G_{lo}. Hence, $\ell(G_{lo}) = \overline{\{t_2t_3t_1t_2, t_3t_1, t_2t_2, t_3t_2t_1t_2t_1t_2, t_3t_2t_1t_1t_2t_2\}}$. The initial condition of G_{lo} will result in the initial condition $M^0_{hi} = [1,0]$ for G_{hi}, where pathways v_1 and v_3 are active. Note that the activation of v_2 is zero since v_2 does not have any boundary transition to fire. Observe that $T_{hi} = \{l_1\}$. G_{mi} is the same as G_{hi} with l_1 replaced by t_3. Thus $\ell(G_{hi}) = \{\varepsilon, t_3\}$. Therefore $\Theta_{\Omega_{mi}}(\ell(G_{lo})) = \{\varepsilon, t_3\} = \ell(G_{mi})$, which means G_{hi} is *consistent* with G_{lo}, with respect to M^0_{lo}. Any string generated by G_{mi} will then have only one t_3, and no more t_3's will appear afterwards. This is consistent with the language generated by G_{lo}. ■

VI. Conclusions

Modeling the cell as a discrete event system helps reduce the complexity of the analysis and synthesis (control) problems, while capturing the essential knowledge. However, due to many factors such as the size of the problem and the uncertainty in the knowledge, the existing techniques for the study of discrete-event system (in particular, supervisory control) need to be extended. In this paper, two algorithms are provided to create high-level and low-level models from the public domain biological knowledge. Furthermore, a method is presented to verify whether or not the obtained abstract model can replace the low-level model (for the purpose of studying pathway sequences) by checking the modeling consistency. These models may be used for developing control strategies for altering cell dynamics.

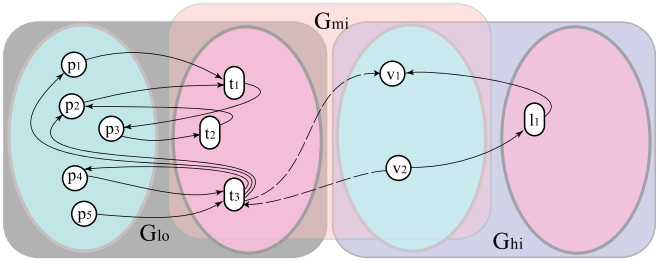


Fig. 5: Petri net example of the high-level and low-level models. The arcs of the middle model are shown using dashed lines

References

[1] E. Sontag, "Molecular systems biology and control," *European Journal of Control*, vol. 11, no. 4, pp. 1 – 40, Jun. 2005.

[2] N. El-Farra, A. Gani, and D. Christofides, "A switched systems approach for the analysis and control of mode transitions in biological networks," *in Proc. of American Contr. Conf.*, vol. 5, pp. 3247 – 3252, Jun. 2005.

[3] A. A. Julius, A. Halasz, V. Kumar, and G. J. Pappas, "Finite state abstraction of a stochastic model of the lactose regulation system of escherichia coli," *in Proc. of 45th IEEE Conf. on Decision and Control*, pp. 19 – 24, Dec. 2006.

[4] D. Degenring, M. Rohl, and A. Uhrmacher, "Discrete event, multi-level simulation of metabolite channeling," *BioSystems*, pp. 29–41, 2004.

[5] M. Chen and R. Hofestädt, "A medical bioinformatics approach for metabolic disorders: biomedical data prediction, modeling, and systematic analysis," *Journal of Biomedical Informatics*, vol. 39, no. 2, pp. 147–159, 2006.

[6] H. Genrich, R. Küffner, and K. Voss, "Executable petri net models for the analysis of metabolic pathways," *International Journal on Software Tools for Technology Transfer (STTT)*, vol. 3, no. 4, pp. 394 – 404, Sep. 2001.

[7] R. Ghosh and C. Tomlin, "Symbolic reachable set computation of piecewise affine hybrid automata and its application to biological modelling: Delta-notch protein signalling," *IEE Proceedings on Systems Biology*, vol. 1, no. 1, pp. 170 – 183, Jun. 2004.

[8] A. Halasz, V. Kumar, M. Imielinski, C. Belta, O. Sokolsky, S. Pathak, and H. Rubin, "Analysis of lactose metabolism in e.coli using reachability analysis of hybrid systems," *IET Systems Biology*, vol. 1, no. 2, pp. 130 – 148, 2007.

[9] M. Poolman, B. Bonde, A. Gevorgyan, H. Patel, and D. Fell, "Challenges to be faced in the reconstruction of metabolic networks from public databases," *IEE Proceedings on Systems Biology*, vol. 153, no. 5, pp. 379 – 384, 2006.

[10] Y. Benenson, B. Gil, U. Ben-Dor, R. Adar, and E. Shapiro, "An autonomous molecular computer for logical control of gene expression," *Nature*, vol. 429, pp. 423 – 429, May 2004.

[11] M. Stanizzi and S. Hall, "Clinical experience with gene therapy for the treatment of prostate cancer," *Reviews in Urology*, vol. 9, pp. S20 – S28, 2007.

[12] S. Hashtrudi-Zad, R. Kwong, and W. Wonham, "Fault diagnosis in discrete-event systems: framework and model reduction," *IEEE Transactions on Automatic Control*, vol. 48, no. 7, pp. 1199 – 1212, Jul. 2003.