Towards Biopathway Modeling and Simulation

Hiroshi Matsuno[1], Sachie Fujita[2], Atsushi Doi[2],
Masao Nagasaki[3], and Satoru Miyano[3]

[1] Faculty of Science, Yamaguchi University,
Yoshida, Yamaguchi 753-8512, Japan
matsuno@sci.yamaguchi-u.ac.jp
[2] Graduate School of Science and Engineering, Yamaguchi University
Yoshida, Yamaguchi 753-8512, Japan
{fujita, atsushi}@ib.sci.yamaguchi-u.ac.jp
[3] Human Genome Center, Institute of Medical Science, University of Tokyo
4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
{masao, miyano}@ims.u-tokyo.ac.jp

Abstract. Petri net has been employed for modeling metabolic pathways as well as signal transduction pathways and gene regulatory networks. The purpose of this paper is to introduce an extension of Petri net called hybrid functional Petri net (HFPN) which allows us to model biopathways naturally and effectively. The method for creating biopathways with HFPN is demonstrated through a well-known biopathway example "*lac* operon gene regulatory mechanism and glycolytic pathway." In order to evaluate this biopathway model, simulations of five mutants of the *lac* operon are carried out by Genomic Object Net which is a biopathway simulator based on the HFPN architecture. The software Genomic Object Net and the HFPN files presented in this paper can be downloaded from http://www.GenomicObject.Net/.

1 Introduction

The notion of Petri net enhanced with its graphical representation creates a high affinity for the understanding of biological systems. It describes a system as a network consisting of places, transitions, and arcs that connect them. Then places contain tokens as their contents, and transitions are active components which can control the contents of places. By assigning biological molecules or stimuli to places and by defining biological reactions as transitions, Petri net can turn to be a model of a biological system. In particular, most biopathways can be abstracted with this notion.

In the literature, Petri net has been employed for modeling metabolic pathways, and qualitative properties of these pathways were discussed [20], where places represent biochemical compounds (metabolites), transitions represent biochemical reactions between metabolites which are usually catalyzed by a certain enzyme, and tokens indicate the presence of compounds. Quantitative simulations by using self-modified Petri net [23] was also made successfully [9], where the number of tokens in a place is used to represent the level of concentrations of the corresponding compound.

W.M.P. van der Aalst and E. Best (Eds.): ICATPN 2003, LNCS 2679, pp. 3–22, 2003.

These Petri nets are discrete in the sense that places can contain nonnegative integers as tokens. On the other hand, it is sometimes more reasonable to regard a biopathway as a hybrid system of discrete and continuous events. We have shown that the gene regulatory network of λ phage can be more naturally modeled as a hybrid system of "discrete" and "continuous" dynamics [17] by employing hybrid Petri net (HPN) [2] and its modification [4]. It is also observed [6] that biopathways can be handled as hybrid systems, e.g. protein concentration dynamics behaves continuously being coupled with discrete switches; protein production is switched on or off depending on the expression of other genes, i.e. presence or absence of other proteins in sufficient concentration.

This paper introduces an extension of hybrid Petri net called hybrid functional Petri net (HFPN) with which we can model biopathways more smoothly. Then we demonstrate how biopathway modeling can be effectively conducted with HFPN by taking, as an example, the *lac* operon gene regulatory mechanism and glycolytic pathway. For this modeling and simulation, we used a software Genomic Object Net (GON) [11,19] that we have developed by realizing the notion of HFPN together with the GUI specially designed for biopathway modeling. HFPN is defined by slightly but inevitably extending HPN so that metabolic pathways, gene regulatory networks and signal transduction pathways may be intuitively modeled.

These extensions and usages are not enough to cover all features in biological systems, but the notion of Petri net would a very promissing architecture for biopathway modeling and simulation, if extensions/modifications could be made so that they reflect the way of thinking of biologists and their intuitions and insights are appreciated in the process of modeling and simulation.

Some practical evidences have been be provided for the usefulness of this notion of HFPN together with the usage of GON. First, in addition to the biopathway modeled in this paper, we have modeled with GON (a) the signal transduction pathway for apoptosis induced by the protein Fas [11], (b) the boundary formation by Notch signaling pathway in *Drosophila* multicellular systems [18], (c) the gene regulatory network of the *Drosophila* circadian rhythms [11], (d) the gene regulatory network of λ phage, and (e) KEGG pathway models are converted to GON XML files for re-modeling and simulation. All these models are available from GON web site [11]. Second, we can give a graphical HFPN description of a system of ordinary differential equations (ODEs) which has been widely used as a representation method in biological simulation systems such as BioDrive [13], Gepasi [15], and E-Cell [22]. In particular, E-Cell can be realized as a subset of GON and we developed a program to convert rule and reactor files of E-Cell to GON XML files except for the C++ programs in E-Cell (this should be done by hand conversion only if necessary) [16]. Third, since HFPN has a graphical representation, our intuitive understanding of the whole system will be enchanced even if the system constitutes a large network of cascades. Generally speaking, in the modeling method based on ODEs, it is difficult to observe the whole system intuitively. On the other hand, GON provides an editor for HFPN which can manipulate biopathway models like pictures. As is emphasized in the design concept of the GUI of BioSPICE [10], the understandability of biopathways is important for modeling and the approach of HFPN with GON

realizes this idea in a similar way. The difference between BioSPICE and GON is that GON equips the engine for simulations. The gene regulatory network of λ phage is also modeled [7] with QSIM simulation environment [12], where LISP is used to describe the model. This is a mathematically beautiful representation but the understandability from the biological viewpoint would remain as another challenge.

In the following section, we define the notion of HFPN in an informal and intuitive way. Then we show how to create the HFPN model of the *lac* operon gene regulatory mechanism and glycolytic pathway in details. This would help readers to model another biopathways of their interests. The model is evaluated with GON by simulation. We deal with various mutants in the *lac* operon for comparison with the wild type.

2 Modeling Methods

2.1 Hybrid Functional Petri Net

This section introduces the notion of *hybrid functional Petri net* (HFPN) which consititues the basic architecture of Genomic Object Net (GON)[1].

Petri net is a network consisting of *place*, *transition*, *arc*, and *token*. A place can hold tokens as its content. A transition has arcs coming from places and arcs going out from the transition to some places. A transition with these arcs defines a *firing rule* in terms of the contents of the places where the arcs are attached.

Hybrid Petri net (HPN) [2] has *discrete place* and *continuous place* as places and *discrete transition* and *continuous transition* as transitions. A discrete place and a discrete transition are the same notions as used in the traditional *discrete Petri net* [21]. A continuous place can hold a nonnegative real number as its content. A continuous transition fires continuously in an HPN and its firing speed is given as a function of values in the places. The graphical notations of a discrete transition, a discrete place, a continuous transition, and a continuous place are shown in Fig. 1, together with three types of arcs. A specific value is assigned to each arc as a weight. When a *normal arc* with weight w is attached to a discrete transition, w tokens are transferred through the normal arc. On the other hand, when a normal arc is attached to a continuous transition, the amount of token that flows is determined by the firing speed of the continuous transition. An *inhibitory arc* with weight w enables the transition to fire only if the content of the place at the source of the arc is less than or equal to w. For example, an inhibitory arc can be used to represent repressive activity in gene regulation. A *test arc* does not consume any content of the place at the source of the arc by firing. For example, a test arc can be used to represent enzyme activity, since the enzyme itself is not consumed.

The discrete Petri net has been extended so that the content of a place can be used as a parameter for the formula describing the weight on the arc from the place that represents the threshold and consumption of tokens for firing [23, 9]. With this extention, biochemical processes are well modeled [9,5,8].

[1] GON is realized with *hybrid functional Petri net with extension* which is enhanced with more types for places (integer, real, boolean, string, vector).

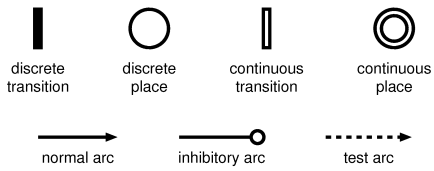

Fig. 1. Basic elements of hybrid (functional) Petri net.

By the definitions of the hybrid Petri net [2] and its modification [4], the firing speed of a continuous transition must be the same as the consuming speed through each arc from its its source place and the contents of all source places are consumed with the same speed. This speed also must be the same as the production speed through each arc from the transition. This is a severe drawback of HPN for biopathway modeling. For example, consider a reaction in which a dimer is cleaved to two monomers.

From these considerations, we define a new kind of transition called a *functional continuous transition* so that any functions can be assigned to arc/transition controling the speed/condition of consumption/production/firing. Then we define HFPN by following the restrictions arising from the hybridity of discrete and continuous features in HPN. The formal definition of HFPN requires a long tedius description and we omit it here. Further necessary details of HFPN will be explained along with the modeling process.

Another feature of HFPN which is not used in the modeling in this paper is hierarchization. The hierarchical representation has been introduced in the hybrid object net [4]. HFPN also inherits this hierarchization capability which is realized in GON. An example of hierarchization can be found in [17].

2.2 *lac* Operon Gene Regulatory Mechanism and Glycolytic Pathway

This section demonstrates how we can model the *lac* operon gene regulatory mechanism and glycolytic pathway in *E. coli* with HFPN by using the biological knowledge in the literature [1,14,24]. The HFPN modeling will start with "transcription control switch" (Fig. 3), then it will be expanded by adding "positive regulation" (Fig. 4), "negative regulation" (Fig. 5), and "hydrolyzing lactose to glucose and galactose" (Fig. 7).

All the parameters in the transitions of the HFPN model are summarized in Tables 1 and 2. We shall show a rough guideline for the usage of transitions in this modeling.

Continuous transition

The firing speed of a continuous transition is given as a simple arithmetic formula such as mX/a, $mX \times mY$, $(mX + mY)/(mX + a \times mY)$, etc., where mX, mY are variables representing the contents of the input places going into

the transition and a is a constant parameter that will be manually tuned. With this speed, the contents of input places will be consumed and simultaneously each output place will receive the amount through the arc.

To each continuous place representing the concentration of some substance, a continuous transition is attached to model the degradation of the substance (e.g. $T1$). The degradation rate is set to be $mX/10$ for high speed degradation (e.g. $T7$) and $mX/10000$ for low speed degradation (e.g. $T8$).

The default weight of the arc going into a continuous transition is set to be "0". If necessary, an appropriate weight is assigned to the arc according to the underlying biological knowledge (e.g. $T67$, $T80$). From the definition, no weight is assigned to the arc outgoing from a continuous transition.

Discrete transition

The default delay time of a discrete transition is set to be "1" and the default weight of an arc from/to a discrete transition is "1". If necessary, the delay time and the weight are appropriately chosen according to the underlying biological knowledge (e.g. $T63$).

Functional continuous transition

A functional continuous transition is used for $T59$ which describes a reaction converting four monomers to one tetramer (see Fig. 6 (c)).

Initial values of places

The places which have initial values greater than zero are listed in Table 3.

2.3 Transcription Control Switch

We will sketch how places, transitions and arcs are determined in the process of modeling.

Fig. 2 shows the structure of the *lac* operon.

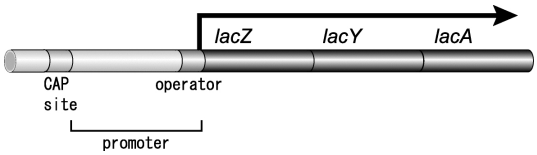

Fig. 2. The *lac* operon: The enzyme β-galactosidase, the product of the *lacZ* gene, hydrolyzes lactose to glucose and galactose. *lacY* encodes the permease that brings lactose into the cell, and *lacA* encodes an acetylase that is believed to detoxify thioglactosides, which, along with lactose, are transported into cells by *lacY*. The "operator" lies within the "promoter", and the "CAP site" lies just upstream of the promoter.

In the absence of lactose, a repressor is bound to the operator. The repressor prevents RNA polymerase from starting RNA synthesis at the promoter. On the

other hand, in the presence of lactose and the absence of glucose, the catabolite activator protein (CAP) is bound to the CAP site. Since the CAP helps RNA polymerase to bind the promoter, the transcription of the *lac* operon begins in this situation.

This regulation mechanism can be expressed by the Petri net of Fig. 3 consisting of only discrete elements.

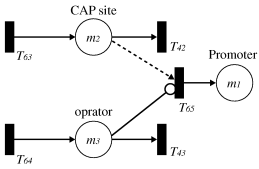

Fig. 3. HFPN representation of the control mechanism of the *lac* operon transcription switch. Only discrete elements are used for representing the switching mechanism.

The place "promoter" ($m1$) represents the status of the transcription of the *lac* operon. That is, if this place contains token(s), the *lac* operon is being transcribed, but if this place has no token, transcription of the operon does not begin. The rates of releasing CAP and repressor from the DNA are assigned to the transitions $T42$ and $T43$ as the delay time, respectively. The production rates of CAP and repressor are also assigned to the transitions $T63$ and $T64$, respectively, as the delay time.

Each time the transition $T65$ fires, the place "promoter" gets one token. This transition fires when both of the following conditions hold; (1) The place "CAP site" ($m2$) contains tokens (this is the case that the protein CAP is bound to the CAP site). (2) The place "operator" ($m3$) has no token (this is the case that the repressor is not bound to the operator).

2.4 Positive Regulation

The DNA binding of CAP is regulated by glucose to ensure that the transcription of the *lac* operon begins only in the absence of glucose. In fact, glucose starvation induces an increase in the intracellular levels of cyclic AMP (cAMP). Transcription of the *lac* operon is activated with the help of CAP to which cAMP binds. When glucose is plentiful, the cAMP level gets down; cAMP therefore dissociates from CAP, which reverts to an inactive form that can no longer bind DNA. This regulatory mechanism by CAP and cAMP is called *positive regulation* of the *lac* operon. Fig. 4 shows an HFPN model which represents positive regulation of the *lac* operon.

Continuous places are used for representing the concentrations of the substances CAP, cAMP, AMP, ADP, and glucose. Tokens in the places "CAP" ($m4$) and "cAMP" ($m5$) should not be consumed by the firing of the transition $T63$,

Table 1. Transitions in Figure 7. All transitions in this figure are listed in "Name" column. The symbol D or C in "Type" column represents the type of transition, discrete transition or continuous transition, respectively. In "Delay/Speed" column, the firing speed of continuous transition or the delay time of discrete transition is described according to the type of transition. The column "From", which represents incoming arc(s) to a transition, is divided into three sub-columns, "variable" (variable name(s) of the place(s) attached to the incoming arc(s)), "weight" (weight of the incoming arc(s)), and "type" (N, T, and I represent normal, test, and inhibitory arcs, respectively). The column "To", which represents outgoing arc(s) from a transition, is divided into two sub-columns, "variable" (variable name(s) of the place(s) attached to the outgoing arc(s)) and "weight" (weight of the outgoing arc(s)).

Name	Type	Delay / Speed	From			To		Comment
			variable	weight	type	variable	weight	
$T1$	C	$m4/10$	$m4$	0	N	—	—	degradation rate of CAP
$T2$	C	$m14/10$	$m14$	0	N	—	—	degradation rate of mRNA repressor
$T3$	C	$m15/10$	$m15$	0	N	—	—	degradation rate of repressor
$T4$	C	$m17/10$	$m17$	0	N	—	—	degradation rate of repressor binding to DNA
$T5$	C	$m18/10$	$m18$	0	N	—	—	degradation rate of repressor not binding to DNA
$T6$	C	$m7/10$	$m7$	0	N	—	—	degradation rate of repressor binding to operator region
$T7$	C	$m19/10$	$m19$	0	N	—	—	degradation rate of *lacZ* mRNA
$T8$	C	$m20/10000$	$m20$	0	N	—	—	degradation rate of LacZ
$T9$	C	$m23/10$	$m23$	0	N	—	—	degradation rate of *lacY* mRNA
$T10$	C	$m24/10000$	$m24$	0	N	—	—	degradation rate of LacY
$T11$	C	$m27/10$	$m27$	0	N	—	—	degradation rate of *lacA* mRNA
$T12$	C	$m28/10000$	$m28$	0	N	—	—	degradation rate of LacA
$T13$	C	$m29/10000$	$m29$	0	N	—	—	degradation rate of lactose outside of a cell
$T14$	C	$m9/10000$	$m9$	0	N	—	—	degradation rate of lactose
$T15$	C	$m8/2$	$m8$	0	N	—	—	degradation rate of arolactose
$T16$	C	$m30/10000$	$m30$	0	N	—	—	degradation rate of galactose
$T17$	C	$m17/10000$	$m17$	0	N	—	—	degradation rate of glucose
$T18$	C	$m10/10000$	$m10$	0	N	—	—	degradation rate of complex
$T19$	C	$m5/10000$	$m5$	0	N	—	—	degradation rate of cAMP
$T20$	C	$m11/10000$	$m11$	0	N	—	—	degradation rate of AMP
$T21$	C	$m12/10000$	$m12$	0	N	—	—	degradation rate of ADP
$T42$	D	1	$m2$	1	N	—	—	CAP releasing rate
$T43$	D	1	$m3$	1	N	—	—	repressor releasing rate
$T45$	C	1	—	—	N	$m4$	—	CAP production rate
$T46$	D	1	—	—	—	$m13$	1	activation of repressor gene
$T57$	D	1.082	$m13$	1	N	$m14$	1	transcription rate of repressor
$T58$	C	$m14$	$m14$	0	N	$m15$	—	translation rate of repressor mRNA
$T59$	C	—	$m15$	—	C†	$m16$	—	conformation rate of repressor
$T60$	C	$96 \times m16/100$	$m16$	0	N	$m7$	—	repressor binding rate to operator
$T61$	C	$399 \times m16/10000$	$m16$	0	N	$m17$	—	repressor binding rate to the DNA other than repressor site
$T62$	C	$m16/10000$	$m16$	0	N	$m18$	—	rate of repressor do not bind any DNA
$T63$	D	1	$m4$ $m5$	1 100	T T	$m2$	1	binding rate of CAP to the CAP site
$T64$	D	1	$m7$ $m8$	1 4	T I	$m3$	1	binding rate of repressor to the operon

† During $m15 > 0$, the amount flow into the transition $T59$ in the speed $(4 \times m15)/5$ from the place "repressor" and the amount flow from the transition $T59$ to the place "4 repressor" in the speed $m15/5$.

Table 2. Transitions in Figure 7 (continued)

Name	Type	Delay / Speed	From			To		Comment
			variable	weight	type	variable	weight	
$T65$	D	1	$m2$	1	T	$m1$	1	logical operation of the places
			$m3$	1	I			"CAP site" and "operator"
$T66$	D	3.075	$m1$	1	T	$m19$	1	transription rate of $lacZ$
						$m21$	1	
$T67$	C	$m19$	$m19$	1	T	$m20$	—	translation rate of $lacZ$
$T68$	D	0.051	$m21$	1	N	$m22$	1	moving rate of RNA polymerase
$T69$	D	1.254	$m22$	1	N	$m23$	1	transription rate of $lacY$
						$m25$	1	
$T70$	C	$m23/2$	$m23$	1	T	$m24$	—	translation rate of $lacY$
$T71$	D	0.065	$m25$	1	N	$m26$	1	moving rate of RNA polymerase
$T72$	D	0.682	$m26$	1	N	$m27$	1	transription rate of $lacA$
$T73$	C	$m27/5$	$m27$	1	T	$m28$	—	translation rate of $lacA$
$T74$	C	$\frac{m24\times m29}{m29+m24\times 10}$	$m29$	0	N	$m9$	—	transforming rate of lactose into a cell
			$m24$	2.5	T			
$T75$	C	$\frac{m20\times m9}{m9+m20\times 10}$	$m9$	0	N	$m30$	—	decomposing rate of lactose to galactose and glucose
			$m20$	5	T	$m6$	—	
$T76$	C	$m9/5$	$m9$	1	T	$m8$	—	producing rate of arolactose from lactose inside of a cell
$T77$	D	0.5	$m3$	1	N	$m10$	1	conformation rate of repressor and arolactose
			$m8$	1	N			
$T78$	C	$m8 \times m16$	$m8$	4	N	$m10$	1	conformation rate of repressor and arolactose
			$m16$	1	N			
$T79$	C	$m5/10$	$m5$	0	N	$m11$	—	reaction rate: cAMP to AMP
$T80$	C	$m11/10$	$m11$	0	N	$m5$	—	reaction rate: AMP to cAMP
			$m6$	5	I			
$T81$	C	$m11/10$	$m11$	0	N	$m12$	—	reaction rate: AMP to ADP
$T82$	C	$m12/10$	$m12$	0	N	$m11$	—	reaction rate: ADP to AMP
$T94$	C	$m29/10$	$m29$	0	T	$m8$	—	producing rate of arolactose from lactose outside of a cell

Table 3. The places having non-zero initial values in Figure 7.

Name	variable	initial value	Comment
CAP	$m4$	5	concentration of CAP
cAMP	$m5$	100	concentration of cAMP
glucose	$m6$	50	concentration of glucose
AMP	$m11$	200	concentration of AMP
ADP	$m12$	200	concentration of ADP
LacZ	$m20$	5	concentration of LacX
LacY	$m24$	2.5	concentration of LacY
LacA	$m28$	1	concentration of LacA
lactose outside of a cell	$m29$	50	lactose outside of a cell

since both of CAP and cAMP are not lost by forming a complex of these two substances. Then, two test arcs are used from the places "CAP" and "cAMP" to the transition $T63$. The weight of the arc from the place cAMP to the transition $T63$ is set to 100, while the weight of the arc from the place "CAP" to that transition is 1, which was determined by manual tuning and referring to the simulation results. After both of the concentrations of CAP and cAMP exceed the thresholds which are given to these two arcs as weights, the transition $T63$ can fire, transferring a token from the transition $T63$ to the place "CAP site."

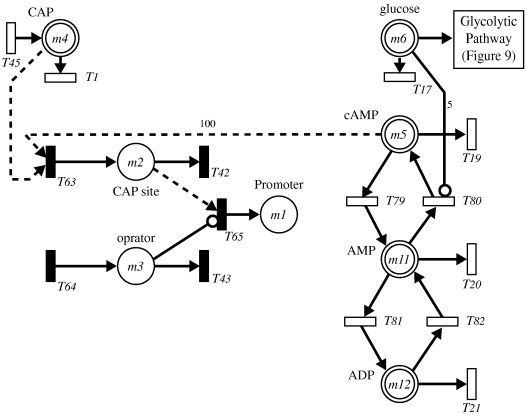

Fig. 4. Positive regulation mechanism: This figure properly contains the HFPN model of Fig. 3.

In general, reactions among cAMP, AMP, and ADP are reversible. The transition $T80$ (the transition $T82$) between the places "cAMP" ($m5$) and "AMP" ($m11$) (the places "AMP" and "ADP" ($m12$)) represents the reversible reaction together with the transition $T79$ (the transition $T81$). To the places "cAMP", "AMP", and "ADP", 1, 200, and 200 are assigned as initial values, respectively.

Recall that when glucose is plentiful, the cAMP level gets down. This phenomenon is represented by the inhibitory arc from the place "glucose" ($m6$) to the transition $T80$. When the concentration in the place "glucose" exceeds the threshold given at this inhibitory arc, the transition $T80$ stops its firing.

In this model, since we suppose that the CAP is produced continuously, this production mechanism is modeled with the place "CAP" and the transitions $T45$ and $T1$. The place "CAP" contains five tokens as an initial value, since CAP is produced by the production mechanism which is independent of the mechanism described here.

Finally, the transitions $T1$, $T17$, $T19$, $T20$, and $T21$ represent the natural degradation of the corresponding substances. Since these transitions are used for representing degradation and do not produce any, no arcs are going out from these transitions.

2.5 Negative Regulation

In the presence of lactose, a small sugar molecule called allolactose is formed in a cell. Allolactose binds to the repressor protein, and when it reaches a high enough concentration, transcription is turned on by decreasing affinity of the repressor protein for the operator site. The repressor protein is the product of the *lacI* gene which is located upstream of the *lac* operon. Actually, after forming a tetramer, the repressor protein can bind the operator site.

By adding this negative regulation mechanism to Fig. 4, we obtain Fig. 5. Since it is known from the literature [14] that repressor should be produced sufficiently prior to the production of other substances, parameters relating to this negative regulation are set faster than other parameters.

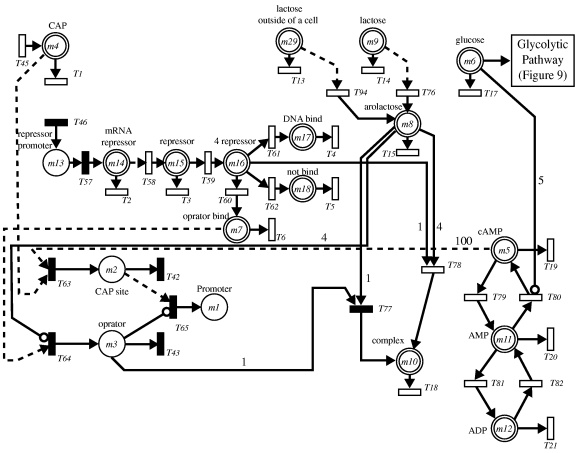

Fig. 5. Negative regulation mechanism: This figure properly contains the HFPN model of Fig. 4.

In our model, a discrete place is used for representing the promoter site of a gene. When the discrete place "repressor promoter"($m13$) gets a token, the transcription of the *lacI* gene begins. We can determine the transcription frequency by the delay rate of the transition $T46$. Transcription and translation mechanisms are modeled by the places "mRNA repressor" ($m14$) and "repressor" ($m15$) and the transitions $T57$, $T2$, $T58$, and $T3$.

The reaction composing a tetramer from monomers (Fig. 6 (a)) can be represented by HPN as is shown in Fig. 6 (b). By comparing this representation with the representation of Fig. 6 (c), we can recognize that HFPN allows us to represent such reaction naturally and intuitively. Tetramer formation is represented in the HFPN by places "repressor" and "4 repressor" ($m16$) and three transitions $T3$, $T59$, and $T60$. Actually, the function $\frac{4\times m15}{5}$ ($\frac{m15}{5}$) is assigned to the input (output) arc to (from) the transition $T59$ as the flow speed. Note that the speed of the input arc is four times faster than the speed of the output arc.

For the repressor forming tetramer, we determined that about 96% of them bind to operator site, about 3.99% of them bind to the other DNA sites, and about 0.01% of them do not bind to DNA. These percentages are determined based on the literature [14]. The places "operator bind" ($m7$), "DNA bind" ($m17$), and "not bind" ($m18$) represent the amount of these repressors. Accord-

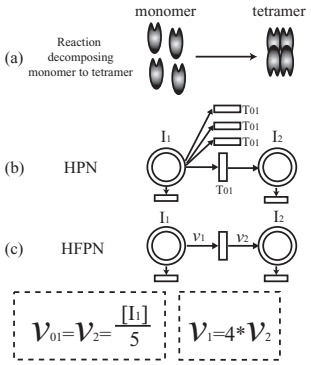

Fig. 6. HPN and HFPN representations of the reaction composing monomers to a tetramer. The speed of v_1 is four times faster than that of v_2. v_{01} is the firing speed of the transition T_{01}. $[\mathrm{I}_1]$ ($[\mathrm{I}_2]$) represents the content of the place I_1 (I_2).

ing to these binding rates, the firing speeds of the transitions $T60$, $T61$, and $T62$ were given by $\frac{96\times m16}{100}$, $\frac{399\times m16}{10000}$, and $\frac{m16}{10000}$, respectively.

We separate the concentration of lactose to two places "lactose outside of a cell" ($m29$) and "lactose" ($m9$) for the convenience of describing the function of the *lacY* gene in the next subsection. The concentration of the arolactose is represented by the place "arolactose" ($m8$) whose accumulation rate is given by the transitions $T94$ and $T76$. It is known [14] that arolactose is produced from the lactose existing outside of a cell as well as the lactose inside a cell. Since it is natural to consider that a production speed of arolactose is faster than a passing rate of arolactose through the cell membrane, the speed of transition $T76$ ($m9/5$) is set to be faster than the speed of transition $T94$ ($m29/10$).

The negative regulation in Fig. 5 is realized in the following way: The place "operator" gets tokens if

- the concentration of the place "operator bind" exceeds the threshold 1 given at the test arc to the transition $T64$ as a weight, and
- the concentration of the place "arolactose" does not exceed the threshold 4 given at the inhibitory arc to the transition $T64$.

Note that the threshold values 1 and 4 are determined according to the observation that four arolactose molecules are required for one tetramer of repressor proteins to bind at the operator site.

The transition $T78$ gives the complex forming late of the arolactose and the tetramer of repressor proteins. The place "complex" ($m10$) represents the concentration of the complex. Arolactose can also release the tetramer of repressor protein from the operator site by forming a complex with it. The transition $T77$ and the arcs from/to the transition realize this mechanism. Discrete transitions are used for the transitions $T77$, since only a discrete transition is available for the arc from the discrete place (continuous amount can not be removed from

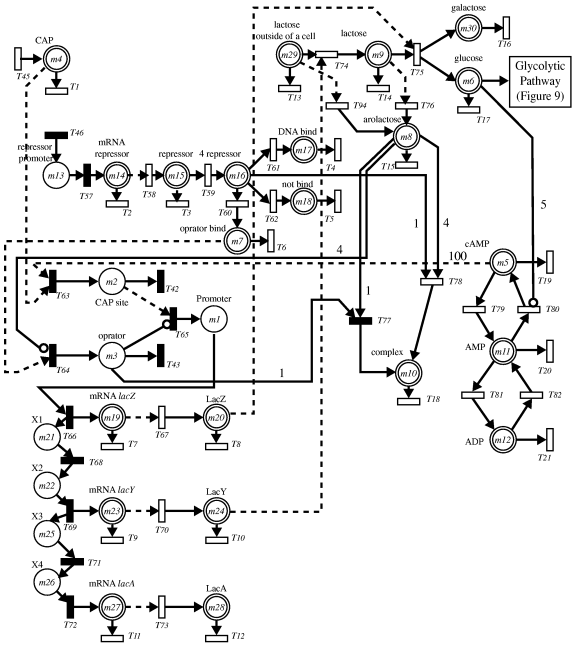

Fig. 7. HFPN modeling of the *lac* operon gene regulatory mechanism. This figure properly contains the HFPN model in Fig. 5.

a discrete place). In order to realize smooth removal of arolactose, small delay time (0.5) is assigned to the transition $T77$.

The transitions $T4$, $T5$, $T6$, $T13$, $T14$, $T15$ and $T18$ represent the natural degradation of the corresponding substances.

2.6 Hydrolyzing Lactose to Glucose and Galactose

The *lac* operon transcription and translation mechanisms are described in Fig. 7 together with the effects of two products of the genes *lacZ* and *lacY* on hydrolyzing lactose to glucose and galactose. The effect of the gene *lacA* is not included in this figure, since the gene *lacA* does not work in both of the *lac* operon gene regulatory mechanism and glycolytic pathway.

The places "mRNA lacZ" ($m19$), "mRNA lacY" ($m23$), and "mRNA lacA" ($m27$) represent the concentrations of mRNAs transcribed from the genes *lacZ*, *lacY*, and *lacA*, respectively. The transcription rates are given at the discrete transitions $T66$, $T69$, and $T72$ as the delay time.

The places "LacZ" ($m20$), "LacY" ($m24$), and "LacA" ($m28$) represent the concentrations of proteins translated from the *lacZ*, *lacY*, and *lacA* mRNAs, and the translation rates are given at the continuous transitions $T67$, $T70$, and $T73$. As is shown in Table 3, we set the initial values of proteins LacZ, LacY, and LacA

as 5, 2.5, 1, respectively. These values are chosen according to the production ratios of LacZ, LacY, and LacA proteins [14]. Actually, the formulas $m19$, $\frac{m23}{2}$, and $\frac{m27}{5}$ are assigned to the transitions $T67$, $T70$, and $T73$ as their speeds, according to the fact that the proteins of *lacZ*, *lacY*, and *lacA* are produced in the ratio $1 : \frac{1}{2} : \frac{1}{5}$.

Degradation rates of mRNAs (proteins) are assigned to the transitions $T7$, $T9$, and $T11$ ($T8$, $T10$, and $T12$).

In this model, to represent the *lac* operon DNA, only discrete elements, discrete transitions $T66$, $T68$, $T69$, $T71$, and $T72$, and discrete places X1 ($m21$), X2 ($m22$), X3 ($m25$), and X4 ($m26$), are used. The discrete places X1 (X3) represents the Boolean status of the transcription of *lacZ* gene (*lacY* gene). That is, each time transcription of *lacZ* (*lacY*) is finished, the place X1 (X3) gets a token. At the discrete transition $T68$ ($T71$), the delay time 0.051 (0.065), which is required for RNA polymerase moving from the end of *lacZ* gene to the beginning of *lacY* gene (the end of *lacY* gene to the beginning of *lacA* gene), is assigned. The delay times 3.075 and 1.254 are assigned to the transitions $T66$ and $T69$, respectively, according to the fact that the length of *lacZ* gene (*lacY* gene) is 3075bp (1254bp). The delay time 0.682 at the transition $T72$ represents the length of *lacA* gene (the length of *lacA* gene is 682bp). The lengths of the genes are obtained from the website [3]. Note that, we can know the transcription status of the gene *lacY* (the gene *lacA*) by observing whether the discrete places X2 (X4) contains token(s) or not.

Recall that the product of the gene *lacZ* is an enzyme which hydrolyzes lactose to glucose and galactose. This reaction is modeled by using the places "lactose", "galactose" ($m30$), and "glucose", and the transitions $T75$, $T16$, and $T17$. In our model, 20 is assigned to each of the places "lactose" and "lactose outside of a cell" as an initial value. Test arc is used from the place "lacZ" to the transition $T75$, since the enzyme is not consumed. We consider that the production rates of glucose and lactose depend on both of the concentration of lactose and the concentration of product of *lacZ* gene. The formula $\frac{m24 \times m29}{m29 + m24 \times 10}$ representing the speed of the transition $T74$ reflects this idea.

We mentioned that the gene *lacY* encodes the permease that brings lactose into the cell. In Fig. 7, this function is realized with the places "LacY", "lactose outside of a cell", and "lactose", and the transitions $T74$, $T13$, and $T14$. Since the product of *lacY* gene is an enzyme, test arc is used from the place "LacY" to the transition $T74$, and the speed of this transition is given by the formula $\frac{m20 \times m9}{m9 + m20 \times 10}$ according to the same idea above.

The weight 2.5 (5) of the arc from the place LacY ($m24$) (LacZ ($m20$)) to the transition $T77$ ($T75$) corresponds to the basal concentration of LacY (LacZ) presented in Table 3.

2.7 Glycolytic Pathway

In the glycolytic pathway, a glucose is converted into two molecules of pyruvate, where the enzymatic reactions shown in Fig. 8 and Table 4 are involved. Fig. 9 shows the HFPN model of the glycolytic pathway.

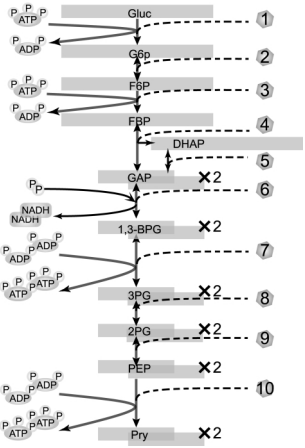

Fig. 8. A part of the glycolytic pathway participating glucose (Gluc), glucose 6-phosphate (G6P), fructose 6-phosphate (F6P), fructose 1,6- diphosphate (FBP), dihydroxyacetone phosphate (DHAP), glyceraldehyde 3-phosphate (GAP), 1,3-diphosphoglycerate (1,3-BPG), 3-phosphoglycerate (3PG), 2-phosphoglycerate (2PG), phosphoenolpyruvate (PEP), and pyruvate (Pry).

Table 4. Mapping between numbers in Fig. 8 and metabolites in reactions

Index	Enzyme / Reaction	Index	Enzyme / Reaction
1	*hexokinase*	2	*phosphoglucose isomerase*
3	*phosphofructokinase*	4	*aldolase*
5	*triosephosphate isomerase*	6	*glyceraldehyde-3-phosphate dehydrogenase*
7	*phosphoglycerate kinase*	8	*phosphoglycerate mutase*
9	*enolase*	10	*pyruvate kinase*
11	*lactate dehydrogenase*		

Main pathway from glucose to pyruvate acid

First we create continuous places corresponding to glucose ($m6$), intermediates ($m31$–$m39$), and pyruvate ($m40$). Default continuous places are introduced at this step though these places are meant to represent the concentrations of the corresponding substrates. Then, by following the pathway in Fig. 8, we put continuous transitions ($T83$–$T93$) together with normal arcs between two consecutive places in the pathway. These transitions and arcs shall represent the reactions, but default transitions and arcs are initially introduced without any parameter tuning. By considering the natural degradation of substrates, we put continuous transitions $T17$ for glucose and $T23$–$T31$ for intermediates with normal arcs. By taking into account the fact that natural degradation is very slow in glycolysis, each of the firing speed of these transitions is given by the formula $mX/10000$ for $X = 17, 23, 24, \ldots, 31$.

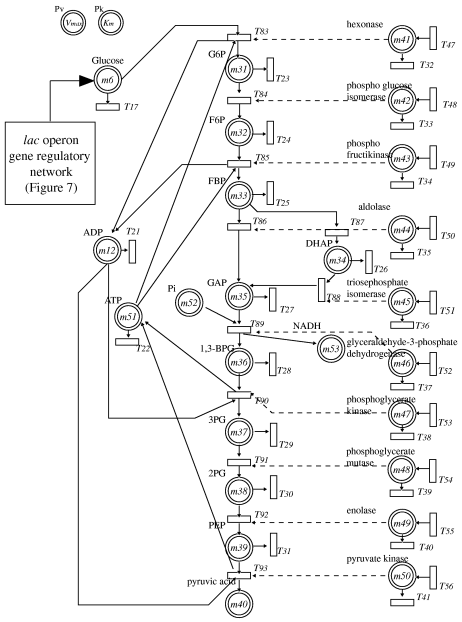

Fig. 9. HFPN model of glycolytic pathway. Michaelis-Menten's equation is applied to the reactions in the main pathway from glucose to pyruvic acid. Test arcs are used to represent enzyme reactions.

Production of ATP from ADP and NADH from phosphoric acid

Next we consider ADP, ATP, Pi (phosphoric acid) and NADH. In the pathway shown in Fig. 8, two ADP molecules and two Pi's are invested to produce four ATP molecules and two NADH molecules. Continuous places "ADP" ($m12$), "ATP" ($m51$), "Pi" ($m52$, initial value=200), and "NADH" ($m53$) are created to represent ADP, ATP, Pi and NADH. We attach continuous transitions $T21$ and $T22$ representing the natural degradation of ADP and ATP. Their firing speeds are set to be very slow ($T21 : m21/10000$, $T22 : m22/10000$) as is set for intermediates in the above. For the process of ATP→ADP in the reaction (1), the normal arc from the place "ATP" to the transition $T83$ and the normal arc from the transition $T83$ to the place "ADP" are introduced. In the same way, normal arcs connected to $T85$ are introduced for the process of ATP→ADP in the reaction (3). Similarly, to represent reactions (7) and (10), transitions $T90$ and $T93$ are used, respectively. Reaction (6) (Pi→NADH) is realized with the transition $T89$. Since ATP and ADP degrade slowly, the formulas $m22/10000$ and $m21/10000$ are assigned to the transitions $T22$ and $T21$ as the degradation speed, respectively.

Installing enzyme reactions

Places having variables $m41$, $m42$,...,$m50$ represent enzyme concentrations. Initial values of these variables are set to 5. The production rates of these enzymes are assigned to the transitions $T47$,$T48$,...,$T56$ whose speeds are set to 1. The transitions $T32$, $T33$,...,$T41$ have the firing speeds $mX/10$'s ($X = 41, 42, ..., 50$), which represent the speed of natural degradation of these enzymes.

Test arcs are used for enzyme reactions, since an enzyme itself is not consumed in the reaction. To each of these test arcs, the value 3 is chosen as a weight of the arc.

Reaction speeds in the main pathway

To represent the reaction speeds in the main pathway, we adopt the Michaelis-Menten equation such as

$$\frac{V_{max}[S]}{K_m + [S]},$$

where $[S]$ is the substance concentration, V_{max} is the maximum reaction speed, and K_m is a Michaelis constant. In our model, we let $V_{max} = 1$ and $K_m = \frac{1}{2}$. The two independent places "Pv" and "Pk" in Fig. 9 represent these variables V_{max} and K_m. The values of V_{max} and K_m can be easily manipulated by changing the contents of the places Pv and Pk, respectively. The Michaelis-Menten equation is used for representing each of the firing speeds of the transitions $T83$, $T84$,..., and $T93$. For example, for the transition $T84$, the formula $\frac{V_{max}m31}{K_m+m31}$ is used.

ADP molecules and two Pi's are invested to produce four ATP molecules and two NADH molecules.

For each of the following two reactions, an HFPN model is created in the following way by assigning different functions to the arcs entering and leaving the transition which represents the reaction:

1. One fructose 1,6-diphosphate (FBP) molecule is invested to produce two glyceraldehyde 3-phosphate (GAP) molecules;
   - "FBP" $\rightarrow T86 : \frac{V_{max}m33}{K_m+m33}$ and
   - $T86 \rightarrow$ "GAP" $: \frac{2\times V_{max}m33}{K_m+m33}$.
2. One dihydroxyacetone phosphate (DHAP) is invested to produce two GAP molecules;
   - "DHAP" $\rightarrow T88 : \frac{V_{max}m34}{K_m+m34}$ and
   - $T88 \rightarrow$ "GAP" $: \frac{2\times V_{max}m34}{K_m+m34}$.

This demonstrates the effectiveness of representing biopathways with HFPN.

3 Evaluation

3.1 Simulation Results of Mutants by GON

The five mutants taken into consideration in this simulation are listed below together with the realization methods of the mutants in the HFPN model of Fig. 8.

lacZ$^-$ a mutant which can not produce β-galactosidase,
- delete the transition $T69$,

lacY$^-$ a mutant which can not produce β-galactoside permease,
- delete the transition $T70$,

lacI$^-$ a mutant in which 4 repressor monomers can not constitute one active repressor tetramer,
- delete the transition $T59$,

lacI^s a mutant to which arolactose can not bind,
- delete the transitions $T77$ and $T78$ together with the inhibitory arc from the place "arolactose" to the transition $T64$,

lacI^{-d} a mutant which can not bind to the DNA,
- delete the transitions $T60$ and $T61$.

Behavior of the wild type

In both of Fig. 10 and Fig. 11, the concentration behaviors of lactose (outside of a cell), lactose, glucose, LacZ (β-galactosidase), and LacY (β-galactoside permease) are shown. From the beginning, glucose is degraded, since it is consumed in the glycolytic pathway. At time point 55, the glucose was consumed, the transcription of the *lac* operon begins, producing the LacZ protein (time=60), and successively, LacY protein begins to be produced. By comparing the concentration behavior of lactose and lactose (outside of a cell), we can recognize that the LacY protein works well. At time point 65, the concentration of LacZ exceeds 10, the decomposition of lactose to glucose and galactose starts, and the concentration of glucose increases again. Just after the glucose is once again completely consumed, the transcription of the *lac* operon is stopped, keeping the concentration of LacZ and LacY proteins at some levels (from the assumption that the degradation speed of these proteins is very show) [1,14].

Behavior of the lac repressor mutant

Fig. 10 shows the simulation results of the mutants *lacI*$^-$, *lacI*^{s}, and *lacI*$^{-d}$ obtained from GON. In the *lacI*$^-$ and *lacI*$^{-d}$ mutants, LacZ protein and LacY protein are produced, while these proteins are not produced in the *lacI*^{s} mutant. Furthermore, in the *lacI*$^-$ and *lacI*$^{-d}$ mutants, the concentrations of LacZ and LacY proteins keep growing (except the period of glucose re-production), even after stopping the decomposition of lactose. Note that these simulation results support the experimental observation [1,14].

Behavior of the lac operon mutant

Fig. 11 shows the simulation results of the mutants *lacZ*$^-$ and *lacY*$^-$ obtained from GON. From this figure, we can observe that, in the *lacZ*$^-$ mutant, once glucose is completely consumed, it is never produced again. On the other hand, in the *lacY*$^-$ mutant, the concentration of lactose (inside of a cell) never grows, since a lactose can not pass a cell membrane in this mutant. Note that these observations from the simulation results also support the experimental observation [1,14].

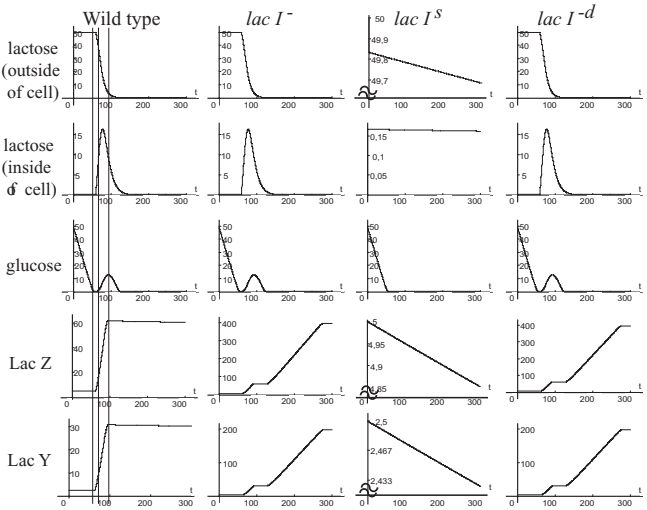

Fig. 10. Simulation results of lac repressor mutant. Concentrations of proteins LacZ and LacY keep growing, since the mutants $lacI^{-}$ and $lacI^{-d}$ lose abilities to bind at the operator site. In the mutant $lacI^{s}$, the transcription of the *lac* operon does not begin, since the repressor can not be removed from the operator site.

4 Conclusion

In this paper, we demonstrated how to build up HFPNs in modeling biopathways as the *lac* operon gene regulatory mechanism and glycolytic pathway as an example. This example was modeled and simulated with GON and the results of five mutants of the *lac* operon gene were shown, which correspond well to the facts described in the literature. Although this result is a well known biological fact, with GON, we have succeeded in discovering one unknown biological phenomenon in multicellular systems [18].

We should emphasize that, essentially, any differential equations can be modeled with HFPN. This means that GON has the potential to simulate biopathway models for other biosimulation tools such as E-Cell and Gepashi. With GON, we have succeeded in modeling some kinds of biopathways [11]. However, at the same time, we have also recognized that the current notion of HFPN is still insufficient to model more sophisticated biopathways including more complex information such as localization, cell interaction, etc. This is one of the reasons to motivate us to develop a new software "Genomic Object Net ver.1.0" (GON 1.0) [19]. GON 1.0 employs the notion of *hybrid functional Petri net with extension* (HFPNe) which allows more "types" for places (integer, real, boolean, string, vector) with which complex information such as localization, etc., can be handled. Furthermore, HFPNe can define a hybrid system of continuous and discrete events together with hierarchization of objects for intuitive creation of complex

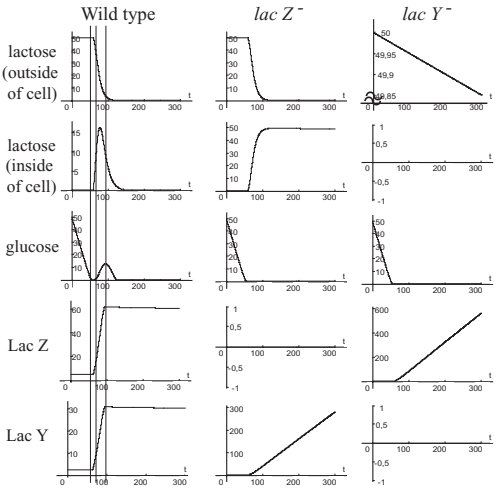

Fig. 11. Simulation results of $lacZ^-$ and $lacY^-$ mutants. We can see that the glucose is not produced in the $lacZ^-$ mutant and the lactose can not enter a cell in the $lacY^-$ mutant.

objects. The news of GON 1.0 including the release date will be announced in the webpage at `http://www.GenomicObject.Net/`.

References

1. Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., Watson, J.: The Molecular Biology of the Cell, Third Edition. Garland Publishing, Inc. (1994)
2. Alla, H., David, R.: Continuous and hybrid petri nets. J. Circuits, Systems, and Computers **8** (1998) 159–188
3. Colibri (*Escherichia coli* genolist browser) `http://genolist.pasteur.fr/Colibri/genome.cgi`.
4. Drath, R.: Hybrid object nets: An object oriented concept for modeling complex hybrid systems. Proc. Hybrid Dynamical Systems, 3rd International Conference on Automation of Mixed Processes, ADPM'98 (1998) 437–442
5. Genrich, H., Küffner, R., Voss, K.: Executable petri net models for the analysis of metabolic pathways. International Journal on Software Tools for Technology Transfer **3** (2001) 394–404
6. Ghosh, R., Tomlin, C.J.: Lateral inhibition through Delta-Notch signaling: a piecewise affine hybrid model. Proc. 4th International Workshop on Hybrid Systems: Computation and Control, Lecture Notes in Computer Science **2034** (2001) 232–246
7. Heidtke, K.R., Schulze-Kremer, S.: Design and implementation of qualitative simulation model of λ phage infection. Bioinformatics **14** (1998) 81–91
8. Heiner, M., Koch, I., Voss, K.: Analysis and simulation of steady states in metabolic pathways with petri nets. Third Workshop and Tutorial on Practical Use of colored Petri Nets and CPN Tools, Universität Aarhus, DAIMI PB–554 (2001) 15–34

9. Hofestädt, R., Thelen, S.: Quantitative modeling of biochemical networks. In Silico Biology **1** (1998) 39-53
10. `http://www.biospice.org`.
11. `http://www.GenomicObject.Net/`.
12. Kuipers, B.: Qualitative reasoning. Modeling and simulation with incomplete knowledge. MIT Press, Cambridge, MA (1994)
13. Kyoda, K., Muraki, M., Kitano, H.: Construction of a generalized simulator for multi-cellular organisms and its application to smad signal transduction. Pacific Symposium on Biocomputing **5** (2000) 317–329
14. Lewin, B.: Genes VI. Oxford University Press and Cell Press. (1997)
15. Mendes, P.: GEPASI: a software for modeling the dynamics, steady states and control of biochemical and other systems. Comput. Appl. Biosci. **9** (1993) 563–571
16. Matsui, M., Doi, A., Matsuno, H., Hirata, Y., Miyano, S.: Biopathway model conversion from E-CELL to Genomic Object Net. Genome Informatics **12** (2001) 290–291
17. Matsuno, H., Doi, A., Nagasaki, M., Miyano, S.: Hybrid petri net representation of gene regulatory network. Pacific Symposium on Biocomputing **5** (2000) 338–349
18. Matsuno, H., Murakami, R., Yamane, R., Yamasaki, N., Fujita, S., Yoshimori, H., Miyano, S.: Boundary formation by notch signaling in *Drosophila* multicellular systems: Experimental observations and a gene network modeling by Genomic Object Net. Pacific Symposium on Biocomputing **8** (2003) 152–163
19. Nagasaki, M., Doi, A., Matsuno, H., Miyano, S.: Genomic Object Net: a platform for modeling and simulating biopathways. submitted for publication.
20. Reddy, V., Mavrovouniotis, M., Liebman, M.: Petri net representations in metabolic pathways. Proc. First ISMB (1993) 328–336
21. Reisig, W.: Petri Nets. Springer-Verlag (1985)
22. Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., Hutchison, C.: E-CELL: software environment for whole cell simulation. Bioinformatics **15** (1999) 72–84
23. Valk, R.: Self-modifying nets, a natural extension of petri nets. Lecture Notes in Computer Science **62** (ICALP '78) (1978) 464–476
24. Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A., Weiner, A.M.: Molecular Biology of the Gene, Fourth Edition. The Benjamin/Cummings Publishing Company Inc (1987)