

# Constructing biological pathway models with hybrid functional Petri nets

Atsushi Doi<sup>1</sup>, Sachie Fujita<sup>1</sup>, Hiroshi Matsuno<sup>1</sup>,  
Masao Nagasaki<sup>2</sup>, and Satoru Miyano<sup>2</sup>

<sup>1</sup> Faculty of Science, Yamaguchi University, Japan

<sup>2</sup> Human Genome Center, Institute of Medical Science, University of Tokyo, Japan

## Abstract

In many research projects on modeling and analyzing biological pathways, the Petri net has been recognized as a promising method for representing biological pathways. From the pioneering works by Reddy *et al.* (1993)[17] and Hofestädt (1994) [9] that model metabolic pathways by traditional Petri net, several enhanced Petri nets such as colored Petri net, stochastic Petri net, and hybrid Petri net have been used for modeling biological phenomena. Recently, Matsuno *et al.* (2003)[14] introduced the hybrid functional Petri net (HFPN) in order to give a more intuitive and natural modeling method for biological pathways than these existing Petri nets. Although the paper demonstrates the effectiveness of HFPN with two examples of gene regulation mechanism for circadian rhythms and apoptosis signaling pathway, there has been no detailed explanation about the method of HFPN construction for these examples. The purpose of this paper is to describe method to construct biological pathways with the HFPN step-by-step. The method is demonstrated by the well-known glycolytic pathway controlled by the *lac* operon gene regulatory mechanism.

## 1 Introduction

A Petri net [18] is method to describe and model concurrent systems. It has been mainly used so far to model artificial systems such as manufacturing systems [16] and communication protocols [23]. The first attempt to use Petri nets for modeling biological pathways was made by Reddy *et al.* [17], giving a method to represent metabolic pathways. Hofestädt expanded this method to model metabolic networks [9]. Subsequently, several enhanced Petri nets have been used to model biological phenomena. Genrich *et al.* [5] modeled metabolic pathways with a colored Petri net by assigning enzymatic reaction speeds to the transitions, and simulated a chain of these reactions quantitatively. Voss *et al.* used the colored Petri net in a different way, accomplishing a qualitative analysis of steady state in metabolic pathways [21]. The stochastic Petri net has been applied to model a variety of biological pathways; the ColE1 plasmid replication [7], the response of the  $\sigma^{32}$  transcription factor to a heat shock [19], and the interaction kinetics of a viral invasion [20]. On the other hand, we have shown that the gene regulatory network of  $\lambda$  phage can be more naturally modeled as a hybrid system of “discrete” and “continuous” dynamics [12] by employing a hybrid Petri net (HPN) architecture [2,3]. It has also been observed [6] that biological pathways can be handled as hybrid systems. For example, protein concentration dynamics, which behave continuously, being coupled with discrete switches. Another example is protein production that is switched on or off depending on the expression of other genes, i.e. the presence or absence of other proteins in sufficient concentrations.

Recently, by extending the notion of HPN, Matsuno *et al.* [14] introduced the hybrid functional Petri net (HFPN) in order to give a more intuitive and natural modeling

method for biological pathways than the existing Petri nets. Although the paper demonstrates the effectiveness of an HFPN with two examples, gene regulation mechanism for circadian rhythms and apoptosis signaling pathway, it only gives the constructed HFPN models for these examples. The purpose of this paper is to demonstrate how biological pathways can be created with an HFPN describing step-by-step the process to construct a model of the *lac* operon gene regulatory mechanism and glycolytic pathway. The constructed HFPN model is verified by simulations of five mutants of *lac* operon on the Genomic Object Net (GON) software package [8, 14]. This software was developed based on the notion of the HFPN together with a GUI specially designed for biological pathway modeling.

## 2 Modeling Biological Pathway with Hybrid Functional Petri Net

In this chapter, first we give a brief definition of the hybrid functional Petri net and give a summary for the modeling method of biological pathways with the hybrid Petri net.

### 2.1 Hybrid functional Petri net: an extended hybrid Petri net for modeling biological reactions

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 two kinds of places *discrete place* and *continuous place* and two kinds of transitions *discrete transition* and *continuous transition*. A discrete place and a discrete transition are the same notions as used in the traditional discrete Petri net [17]. A continuous place can hold a nonnegative real number as its content. A continuous transition fires continuously at the speed of a parameter assigned at the continuous transition. 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* is attached to a discrete/continuous transition,  $w$  tokens are transferred through the normal arc, in either of normal arcs coming from places or going out to places. 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.

*Hybrid dynamic net* (HDN) [3] has a similar structure to the HPN, using the same kinds of places and transitions as the HPN. The main difference between HPN and HDN is the firing rule of continuous transition. As we can know from the description about HPN above, for a continuous transition of HPN, the different amounts of tokens can be flowed

through the two types of arcs, coming from/going out the continuous transition. In contrast, the definition of HDN does not allow to transfer different amount through these two types of arcs. However, HDN has the following firing feature of continuous transition which HPN does not have; ``the speed of continuous transition of HDN can be given as a function of values in the places''.

From the above discussion, we can know that each of HPN and HDN has its own feature for the firing mechanism of continuous transition. As a matter of fact, both of these features of HPN and HDN are essentially required for modeling common biological reactions. (See the example of Fig. 6 in which four monomers compose one tetramer.)

This motivated us to propose hybrid functional Petri net (HFPN) [14] which includes both of these features of HPN and HDN. Moreover, HFPN has the third feature for arcs, that is, a function of values of the places can be assigned to any arc. This feature was originated from the idea in the paper [10] which was introduced in order to realize the calculation of dynamic biological catalytic process on Petri net based biological pathway modeling. In fact, these three features are realized in HFPN by introducing a new type transition called a *functional continuous transition* which allows us to assign any functions to arcs and transitions for controlling the speed/condition of consumption/production/firing. An example to use the functional continuous transition is given in the beginning of Section 3.

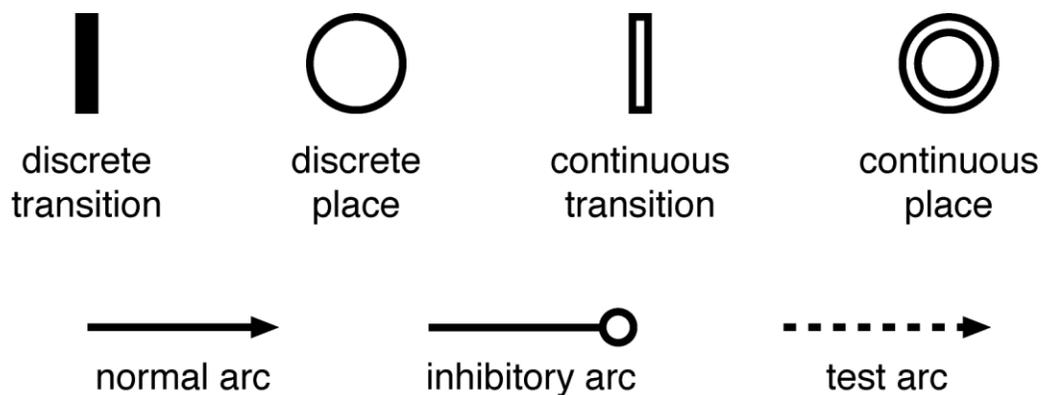


Fig. 1: Basic elements of HPN, HDN, and HFPN.

## 2.2 Usage of discrete and continuous elements

Biological pathways essentially consist of discrete parts such as a genetic switch control and continuous parts such as a metabolic reaction. These discrete and continuous parts can be represented by discrete elements (discrete place and discrete transition) and continuous elements (continuous place and continuous transition) of HFPN.

For example, a control system turning on or off the gene expression with operator site can be represented by discrete elements. That is, if the discrete place has a token, the protein necessary for activating the operator site has bound to the operator, that means the

gene expression turn on. In addition, using the delay concept of the discrete transition, we can easily describe the transcription which happens after a certain time.

On the other hand, biological phenomenon such as transcription, translation, and enzymic and metabolic reactions have been treated as events whose conditions change continuously. By modeling the transcription and the translation with continuous elements in the following way, expression levels of mRNA and proteins can be simulated. Continuous places are used for storing concentrations of mRNA and protein. Reaction speeds of transcription and translation are assigned at parameters of continuous transitions. In addition, common formulas for biological reactions such as Michaelis-Menten's equation can be modeled almost directly with assigning concentrations of substrate and product to continuous places and the formula for the reaction of them to the continuous transition between these two places.

### 3 HFPN modeling of the *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 an HFPN. Biological facts used for constructing this model are described in the biological literature [1], [11], [22]. The HFPN modeling will start with a "transcription control switch" (Fig. 3), then it will be enhanced gradually by adding "positive regulation" (Fig. 4), "negative regulation" (Fig. 5), and "hydrolyzing lactose to glucose and galactose" (Fig. 7). This step-by-step explanation based on the well-known biological facts helps readers to understand how HFPNs are created according to the biological knowledge.

All the parameters in the transitions of the HFPN model are summarized in Tables 1 and 2. In this example, we shall show a rough guideline for usage of continuous and discrete transitions. In accordance with the guideline below, readers can understand the modeling manner of biological pathway with HFPN described in the subsections from 3.1 to 3.4. For the detail explanation about firing rules, refer to the paper [14].

#### *Continuous transition*

The firing speed of a continuous transition is described by a simple arithmetic formula such as  $mX/a$ ,  $mX \times mY$ ,  $(mX \times mY)/(mX + a \times mY)$ , etc., where  $mX$ ,  $mY$  are variables representing the contents of the input places going into the transition, and  $a$  is a constant parameter to be tuned manually. With this speed, the contents of input places will be consumed and simultaneously each output place will receive the corresponding amount through the arc. For example, there is a description about the reaction of the protein LacZ, which says that the degradation speed of LacZ is much slower than its production rate. Based on this description (biological knowledge), we used the formula  $m20/10000$  as the degradation speed (transition *T8*) and  $m19$  as the production speed of LacZ. Basically, all transition speeds are given in the same way. That is, in our modeling method, a speed of biological reaction is defined relatively without a unit.

To each continuous place representing the concentration of some substance, a continuous transition is attached to model the degradation of the substance. The degradation rate is set to be  $mX/10$  for high speed degradation (e.g.  $T 1$ ) and  $mX/10000$  for low speed degradation (e.g.  $T 8$ ). Following the definition, no weight is assigned to the arc going out from a continuous transition.

The default weight of the arc going into a continuous transition is set to be zero. If necessary, an appropriate weight is assigned to the arc according to the underlying biological knowledge.

(See the explanation in the subsection 3.4 for the transitions  $T67$ ,  $T70$ ,  $T73$ ).

#### *Discrete transition*

The default delay time of a discrete transition is set to be one and the default weight of an arc from/to a discrete transition is one. If possible, the delay time and the weight are appropriately chosen according to the underlying biological knowledge (See the explanation in the subsection 3.4 for the transitions  $T 66$ — $T73$ ).

#### *Functional continuous transition*

A functional continuous transition is used for  $T 59$ , which describes a reaction converting four monomers to one tetramer (Fig.6 (a)). Recall that HPN can not have functions as parameters of the transitions. We can not model this type of reaction with HPN, since, a reaction speed of this type is always assigned to a transition as a function of values of the places as shown in the above. In contrast, it is possible to model this reaction with the HDN as shown in Fig.6 (b). However, as is shown in Fig.6 (b), the constructed HDN model is not natural or intuitive. Recall that a functional continuous transition of HFPN allows us to assign any functions to arcs and transitions for controlling the speed/condition of consumption, production, or firing. From Fig. 6 (c), we can see that this reaction can be modeled naturally and intuitively with a functional continuous transition.

At an arc of HFPN, two kinds of parameters “threshold” and “speed” are assigned. The continuous transition attached at the head of the arc can not fire unless the content of the place attached at the tail of the arc exceeds the “threshold” of the arc. On the other hand, “speed” is a value or a function, defining the amount flowing through the arc.

#### *Initial values of places*

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

### **3.1 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.

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 to the promoter, the transcription of the *lac* operon can begin. This regulation mechanism can be expressed by the HFPN of Fig. 3, consisting of only discrete elements. The place “promoter” ( $m_1$ ) represents the status of the transcription of the *lac* operon. If this place contains token(s), the *lac* operon is being transcribed, whereas if this place contains no tokens, transcription of the operon does not begin. The rates of releasing CAP and repressor from the DNA are assigned to the transitions  $T_{42}$  and  $T_{43}$  as the delay times, respectively. The production rates of CAP and repressor are assigned to the transitions  $T_{63}$  and  $T_{64}$ , respectively, as the delay times. Each time the transition  $T_{65}$  fires, the place “promoter” receives one token. This transition fires when both of the following conditions hold: (1) The place “CAP site” ( $m_2$ ) contains tokens (this is the case in which the protein CAP is bound to the CAP site). (2) The place “operator” ( $m_3$ ) has no token (this is the case in which the repressor is not bound to the operator).

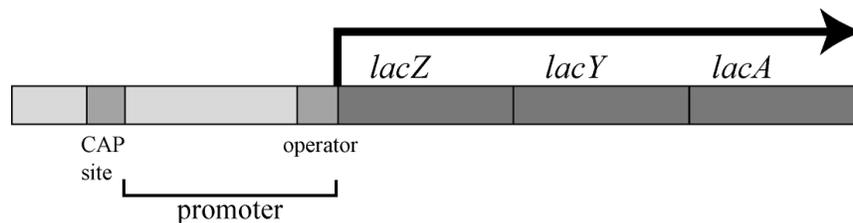


Fig. 2: The *lac* operon: The enzyme  $\beta$ -galactosidase, produced by the *lacZ* gene, hydrolyzes lactose to glucose and galactose. *lacY* encodes the permease that brings lactose into the cell, and *lacA* encodes an acetylase that is believed to detoxify thiogalactosides, which, along with lactose,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

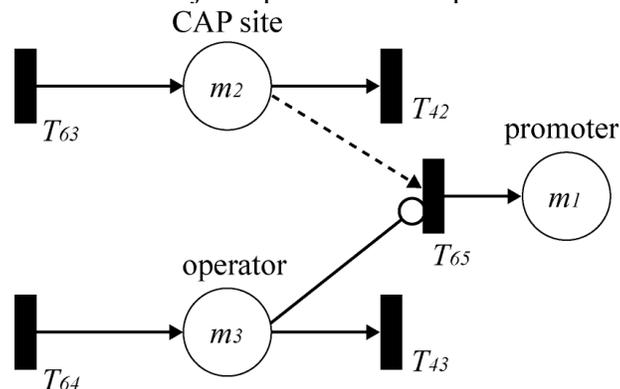


Fig. 3: The HFPN representation of the control mechanism of the *lac* operon transcription switch. Only discrete elements are used for representing the switching mechanism.

### 3.2 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, cAMP levels drop; 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 representing 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  $T 63$ , since both CAP and cAMP are not lost when forming a complex. Accordingly, two test arcs are used from the places “CAP” and “cAMP” to the transition  $T 63$ . The weight of the arc from the place cAMP to the transition  $T 63$  is set to 100, while the weight of the arc from the place “CAP” is 1, which was determined by manual tuning and referring to the simulation results. After the concentrations both of CAP and cAMP exceed the thresholds which are given to these two arcs as weights, the transition  $T 63$  can fire, transferring a token from the transition  $T 63$  to the place “CAP site.”

In general, reactions among cAMP, AMP, and ADP are reversed. The transition  $T 80$  (the transition  $T 82$ ) between the places “cAMP” ( $m5$ ) and “AMP” ( $m11$ ) (the places “AMP” and “ADP” ( $m12$ )) represents the reversible reaction together with the transition  $T 79$  (the transition  $T 81$ ). To the places “cAMP”, “AMP”, and “ADP”, 1, 200, and 200 are assigned as initial values, respectively.

Recall that when glucose is plentiful, cAMP levels drop. This phenomenon is represented by the inhibitory arc from the place “glucose” ( $m6$ ) to the transition  $T 80$ . When the concentration in the place “glucose” exceeds the threshold given at this inhibitory arc, the transition  $T 80$  stops firing.

In this model, since we suppose that CAP is produced continuously, its production mechanism is modeled with the place “CAP” and the transitions  $T45$  and  $T1$ . The initial amount of the place “CAP” is set to 5, since CAP is produced by a production mechanism independent of the mechanism being described here.

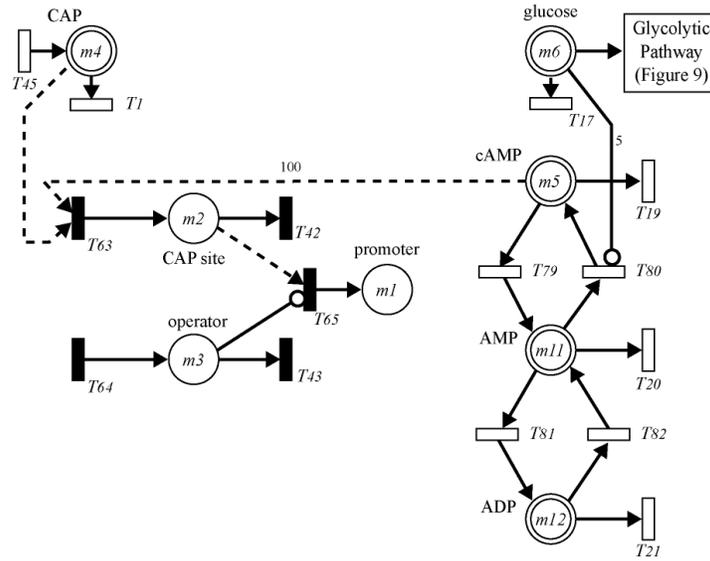


Fig. 4: Positive regulation mechanism: This figure contains the HFPN model of Fig. 3.

Finally, the transitions  $T1$ ,  $T17$ ,  $T19$ ,  $T20$ , and  $T21$  represent the natural degradation of the corresponding substances. Since these transitions represent only degradation and no production, no arcs are going out from these transitions.

### 3.3 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 sufficiently high concentration, transcription is turned on by decreasing the 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 to the operator site.

By adding this negative regulation mechanism to Fig. 4, Fig. 5 is obtained. Since it is known from the literature [11] that repressor should be produced sufficiently prior to the production of other substances, parameters relating to this negative regulation are set to faster values than other parameters. In our model, a discrete place is used for representing the promoter site of a gene. When the discrete place “repressor promoter”(m13) receives 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 forming a tetramer from monomers (Fig. 6 (a)) can be represented by the HDN as is shown in Fig. 6 (b). (HPN can not model this type of reaction, since, any

reaction speed should be realized by assigning a function of values of the places to the continuous transition.) By comparing this representation and the representation of Fig. 6 (c), we recognize that an 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  $T60$ . The function  $\frac{4 \times m15}{5} \left( \frac{m15}{5} \right)$  is assigned for

the input (output) arcs to (from) the transition  $T59$  as a 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 the 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 from the description in the literature [11]. The places “operator bind” ( $m7$ ), “DNA bind” ( $m17$ ), and “not bind” ( $m18$ ) represent the amount of these repressors. According 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. Concentration of the allolactose is represented by the place “allolactose” ( $m8$ ) whose accumulation rates are given at the transitions  $T94$  and  $T76$ . It is known [9] that allolactose 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 allolactose is faster than the passing rate of allolactose 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 was realized in the following way:

- The place “operator” receives 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 “allolactose” does not exceed the threshold 4 given at the inhibitory arc to the transition  $T64$ .

Four molecules of allolactose need to bind with one molecule of tetramer repressor. Accordingly the function of input arc from the place “allolactose” to the transition  $T78$  should be four times faster than that of input arc from the place “4 repressor”. In summary, we gave formulas,

- $4 \times m8 \times m16$  to the arc from the place “allolactose” to  $T78$ ,
- $m8 \times m16$  to the arc from the place “4 repressor” to  $T78$ , and
- $m8 \times m16$  to given at the arc from the  $T78$  to the place “complex”.



the places to the continuous transition.

The transition  $T78$  gives the complex forming rate of the allolactose and the tetramer of repressor proteins. The place “complex” ( $m10$ ) represents the concentration of the complex. Allolactose can also release the tetramer of repressor proteins 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 (a continuous amount can not be removed from a discrete place). In order to realize a smooth removal of allolactose, a 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.

### 3.4 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 *lacA* gene does not play a role in either the *lac* operon gene regulatory mechanism or the 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, and the transcription rates are given at the discrete transitions  $T66$ ,  $T69$ , and  $T72$  as the delay times.

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 to 5, 2.5, 1, respectively. These values are chosen according to the production ratios of LacZ, LacY, and LacA proteins [11]. Actually, the formulas  $m19$ ,  $\frac{m32}{2}$ , and  $\frac{m27}{5}$  are assigned to the transitions  $T67$ ,  $T70$ , and  $T73$  as the 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) receives a token. At the discrete transition  $T68$  ( $T71$ ), the delay time 0.051 (0.065), which is required for the RNA polymerase moving from the end

of the *lacZ* gene to the beginning of the *lacY* gene (the end of the *lacY* gene to the beginning of the *lacA* gene), is assigned. The delay times 3.075 and 1.254 are assigned to the transitions  $T 66$  and  $T 69$ , respectively, according to the fact that the length of the *lacZ* gene (*lacY* gene) is 3075bp (1254bp). The delay time 0.682 at the transition  $T 72$  represents the length of the *lacA* gene (the length of the *lacA* gene is 682bp). The lengths of the genes are obtained from the website [4]. Note that we can know the transcription status of the gene *lacY* (the gene *lacA*) by observing whether the discrete places  $X2$  ( $X4$ ) contains tokens 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  $T 75$ ,  $T 16$ , and  $T 17$ . In our model, 20 is assigned to each of the places “lactose” and “lactose outside of a cell” as an initial value. A test arc is used from the place “lacZ” to the transition  $T 75$ , since the enzyme is not consumed.

We consider that the production rates of glucose and lactose depend on both the concentration of lactose and the concentration of the product of the *lacZ* gene. The formula

$\frac{m24 \times m29}{m29 + m24 \times 10}$  representing the speed of the transition  $T 74$  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  $T 74$ ,  $T 13$ , and  $T 14$ . Since the product of *lacY* gene is an enzyme, test arc is used from the place “LacY” to the transition  $T 74$ , and the speed of this

transition is given by the formula  $\frac{m20 \times m9}{m9 + m20 \times 10}$  similarly above.

The weight 2.5 (5) of the arc from the place LacY ( $m24$ ) (LacZ ( $m20$ )) to the transition  $T 74$  ( $T 75$ ) corresponds to the basal concentration of LacY (LacZ) presented in Table 3.



### 3.5 Glycolytic Pathway

In the glycolytic pathway, glucose is converted into two molecules of pyruvate. The enzymatic reactions shown in Fig. 8 and Table 4 are involved in this reaction.

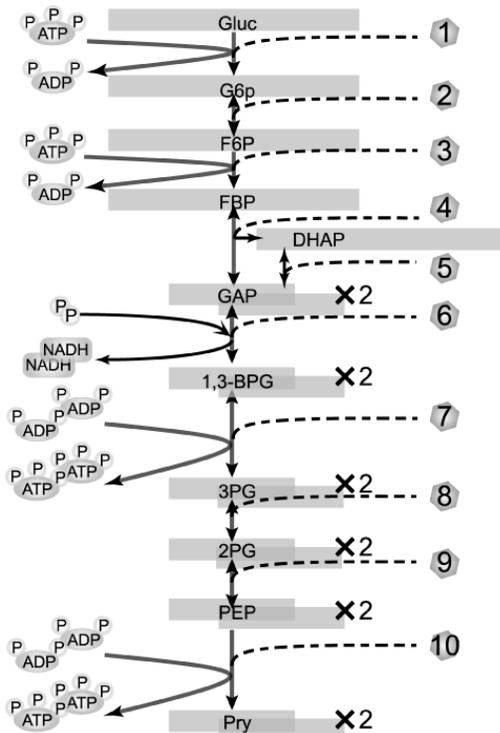


Fig. 8: A part of the glycolytic pathway consisting of glucose (Gluc), glucose 6-phosphate (G6P), fructose 6-phosphate (F6P), fructose 1,6-diphosphate (FBP), dihydroxyacetone phosphate (DHAP), glyceraldehyde 3-phosphate (GAP), 1,3-diphosphoglycerate (1,3-BPG), 3-phosphoglycerate (3PG), 2-phosphoglycerate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

More details about this pathway is explained in more detail in the following modeling process. Fig. 9 shows the HFPN model of the glycolytic pathway.

#### *Main pathway from glucose to pyruvate acid*

First we create continuous places corresponding to glucose ( $m_6$ ), intermediates ( $m_{31}$ – $m_{39}$ ), and pyruvate ( $m_{40}$ ). 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 ( $T_{83}$ – $T_{92}$ ) 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  $T 17$  for glucose and  $T 23$ – $T 31$  for intermediates with normal arcs. By taking into account the fact that natural degradation is very slow in glycolysis, the firing speed of these transitions is given by the formula  $mX/10000$ , where  $X = 17, 23, 24, \dots, 31$ .

*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 the continuous transitions  $T 21$  and  $T 22$  representing the natural degradation of ADP and ATP. Their firing speeds are set to be very slow ( $T 21 : m21/10000$ ,  $T 22 : m22/10000$ ), similarly as for the intermediates above. For the process of ATP→ADP in the reaction (1), the normal arc from the place "ATP" to the transition  $T 83$  and the normal arc from the transition  $T 83$  to the place "ADP" are introduced. In the same way, normal arcs connected to  $T 85$  are introduced for the process of ATP→ADP in the reaction (3). Similarly, to represent reactions (7) and (10), transitions  $T 89$  and  $T 92$  are used, respectively. Reaction (6) (Pi→NADH) is represented by the transition  $T 88$ . Since we consider ATP and ADP to degrade slowly, the formulas  $m22/10000$  and  $m21/10000$  are assigned to the transitions  $T 22$  and  $T 21$  as degradation speed, respectively.

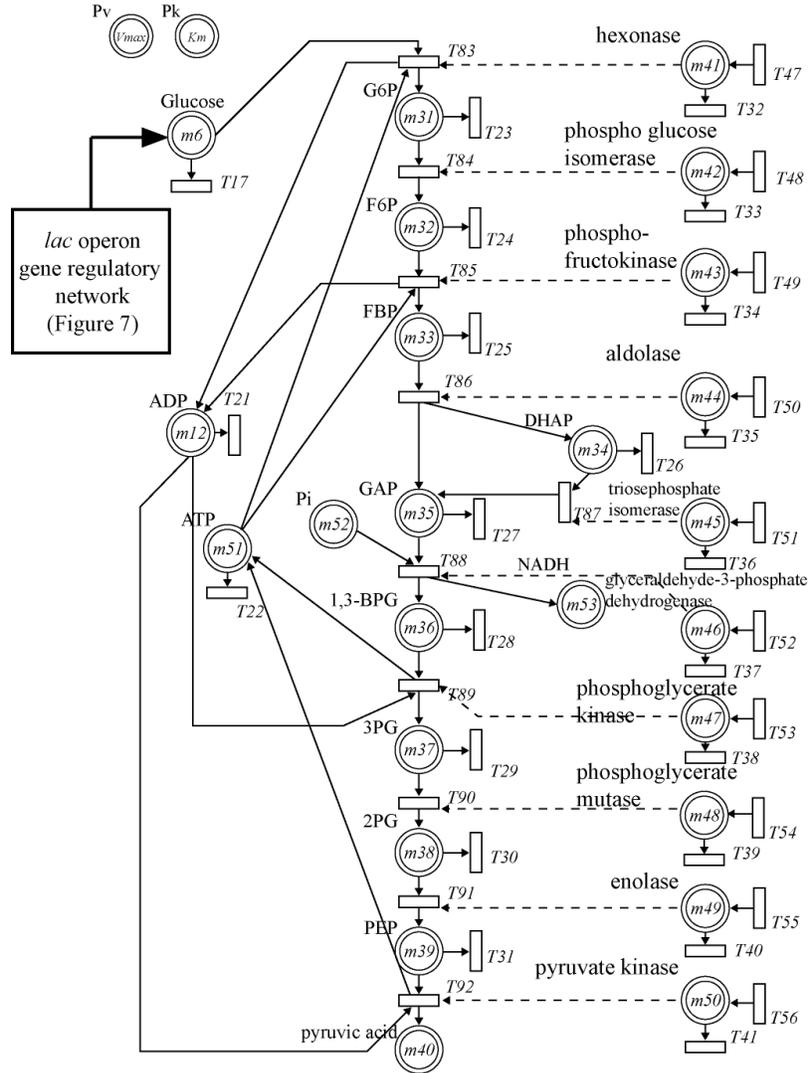


Fig. 9: HFPN model of glycolytic pathway. Michaelis-Menten's equation is applied to the reactions in the main pathway from glucose to pyruvic acid. Test arcs are used to represent enzyme reactions.

#### Including enzyme reactions

Places having variables  $m_{41}, m_{42}, \dots, m_{50}$  represent enzyme concentrations. Initial values of these variables are set to 5. The production rates of these enzymes are assigned to the transitions  $T_{47}, T_{48}, \dots, T_{56}$  whose speeds are set to 1. The transitions  $T_{32}, T_{33}, \dots, T_{41}$  have the firing speeds  $m_X/10$  ( $X = 41, 42, \dots, 50$ ), 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:

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

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  $K_m = 1$ . 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 easily be 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  $T_{83}$ ,  $T_{84}$ , ..., and  $T_{92}$ . For example, for the transition  $T_{84}$ , the formula

$$\frac{V_{\max}m_{31}}{K_m + m_{31}}$$
 is used.

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

## **4 Simulations by Genomic Object Net**

In the previous section, we demonstrated how a biological pathway is modeled with HFPN through an example of the *lac* operon gene regulatory mechanism and glycolytic pathway. This section verifies the HFPN model by analyzing simulation results obtained from GON. GON is the software based on the notion of the HFPN and has a GUI specially designed for biological pathway modeling. GON was developed in Java so that it can run on any platforms. The modeled biological pathways is saved as an original XML format. One of the interesting function of GON is that any place or transition of HFPN can be replaced to an image having biological meaning. Fig. 10 is a screen shot of GON after applying this function several times. In this screen shot, for example, the discrete places “CAP site”, “operator”, and “promoter” in Fig.3 were replaced to the corresponding images.



### Behavior of the wild type

In both of Figures 11 and 12, the concentration behaviors of lactose (outside of a cell), lactose, glucose, LacZ ( $\beta$ -galactosidase), and LacY ( $\beta$ -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 *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 know that the LacY protein works well. At time point 65 the concentration of LacZ exceeds 10, decomposition of lactose to glucose and galactose starts, increasing the concentration of glucose 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 slow) [1], [11].

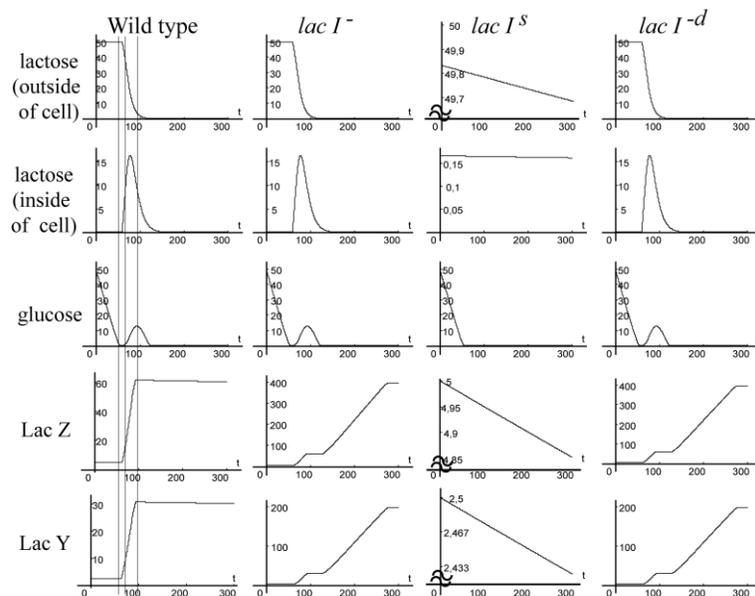


Fig. 11: Simulation results of the *lac* repressor mutant. Concentrations of proteins LacZ and LacY keep growing, since the mutants *lacI* and *lacI<sup>d</sup>* lose the ability to bind at the operator site. In the mutant *lacI<sup>s</sup>*, the transcription of the *lac* operon does not begin, since the repressor can not be removed from the operator site.

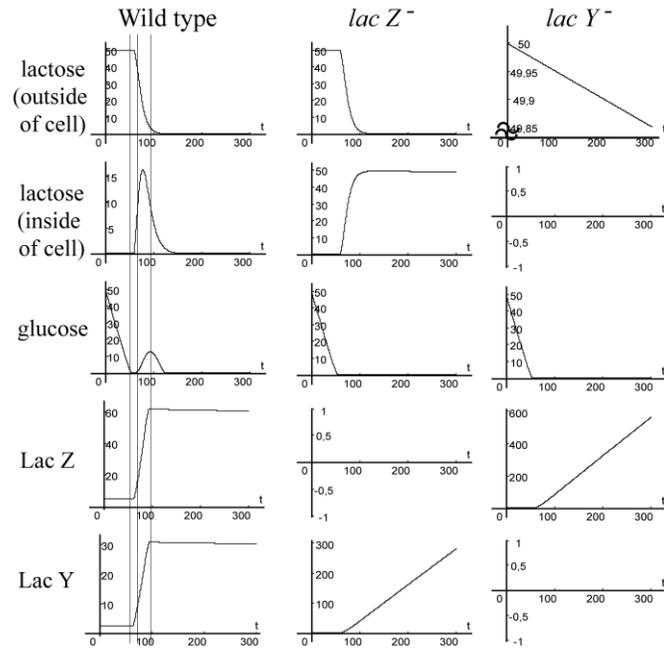


Fig. 12: Simulation results of *lacZ* and *lacY* mutants. We can see that the glucose is not produced in the *lacZ* mutant and the lactose can not enter a cell in the *lacY* mutant.

#### *Behavior of the lac repressor mutant*

Fig. 11 shows the simulation results of the mutants *lacI*, *lacI<sup>s</sup>*, and *lacI<sup>d</sup>* obtained from GON. In the *lacI* and *lacI<sup>d</sup>* mutants, LacZ protein and LacY protein are produced, while these proteins are not produced in the *lacI<sup>s</sup>* mutant. Furthermore, in the *lacI* and *lacI<sup>d</sup>* mutants, the concentrations of LacZ and LacY proteins keep growing (except during the period of glucose re-production), even after the decomposition of lactose has ended. Note that these simulation results support the biological experimental results [1], [11].

#### *Behavior of the lac operon mutant*

Fig. 12 shows the simulation results of the mutants *lacZ* and *lacY* obtained from GON. From this figure, we can observe that, in the *lacZ* mutant, glucose is never produced again once it is completely consumed. On the other hand, in the *lacY* mutant, the concentration of lactose (inside the cell) never grows, since lactose can not pass a cell membrane in this mutant. Note that these observations from the simulation results support the experimental results [1], [11].

## 5 Conclusion

In this paper, we demonstrated how to construct HFPNs in modeling biological pathways

with the example of the *lac* operon gene regulatory mechanism and glycolytic pathway. This example was simulated by GON and the results of five mutants of *lac* operon gene were shown, which correspond well to the facts described in the literature. The purpose of this paper was to describe method to construct biological pathways with the HFPN. Therefore we introduced the well-known glycolytic pathway controlled by the *lac* operon gene regulatory mechanism.

Although this paper deals with a known biological phenomenon, with GON, we succeeded in discovering one unknown biological phenomenon in multicellular systems [13]. We analyzed the mechanism of Notch-dependent boundary formation in the *Drosophila* large intestine, by experimental manipulation of Delta expression and computational modeling and simulation by GON. Boundary formation representing the situation in the normal large intestine was shown in the simulation. By manipulating Delta expression in the large intestine, a few types of disorder in boundary cell differentiation were observed, and similar abnormal patterns were generated by the simulation. Simulation results suggest that parameter values representing the strength of cell-autonomous suppression of Notch signaling by Delta are essential for generating two different modes of patterning: lateral inhibition and boundary formation, which could explain how a common gene regulatory network results in two different patterning modes in vivo.

We should emphasize that, essentially, any type of differential equations can be modeled with HFPN. This means that GON has the potential to simulate biological pathway models for other biosimulation tools such as E-Cell and Gepasi. More concretely, if the users have a skill of programming, they can develop original functions as the Java class files, which can be called from the transitions or the arcs of HFPN. This means that, any function of E-Cell or Gepasi can be included as Java class file in GON.

With GON, we have succeeded in modeling many kinds of biological pathways [8]. However, at the same time, we recognized that the current notion of HFPN is still insufficient to model more sophisticated biological pathways, including more complex information such as localization, cell interaction, etc.

This is one of the reasons to motivate us to create a “hybrid functional Petri net with extension (HFPNe)” by enhancing the concept of HFPN [15]. The HFPNe allows more “types” for places (integer, real, boolean, string, vector) with which complex information can be handled. In other words, the definition of these types allow us to treat other existing Petri nets as subsets of the HFPNe. For example, a traditional Petri net (with only discrete elements) is treated as the HFPNe using only “integer type”. Furthermore, HFPNe can define a hybrid system of continuous and discrete events together with a hierarchization of objects for an intuitive creation of complex objects. Genomic Object Net is developed based on the notion of an HFPNe. The biological pathway models constructed with the HFPNe will be reported in the near future.

## References

- [1] Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson J. (1994). *The Molecular Biology of the Cell, Third Edition*. Garland Publishing, Inc.

[2] Alla, H. and David, R. (1998). Continuous and hybrid Petri nets. *Journal of Circuits, Systems, and Computers* 8, 159–188.

[3] Drath, R. (1998). 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*, 437–442.

[4] *Escherichia coli* genolist browser. <http://genolist.pasteur.fr/Colibri/genome.cgi> .

[5] Genrich, H., Küffner, R. and Voss, K. (2001). Executable Petri net models for the analysis of metabolic pathways. *International Journal on Software Tools for Technology Transfer*, **3(4)**, 394–404.

[6] Ghosh, R. and Tomlin, C. (2001). 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*, 232–246.

[7] Goss, P.J.E. and Peccoud, J. (1998). Quantitative modeling of stochastic systems in molecular biology by using Stochastic Petri nets. *Proc. Natl. Acad. Sci. USA* 95, 6750-6755.

[8] <http://www.GenomicObject.Net/>.

[9] Hofestädt, R. (1994). A Petri net application to model metabolic processes. *SAMS* 16, 113-122.

[10] Hofestaedt R. and Thelen, S. (1998). Quantitative modeling of biochemical networks, *In Silico Biology* 1, 0006, <http://www.bioinfo.de/isb/1998/01/0006/>.

[11] Lewin, B. (1997). *Genes VI*. Oxford University Press and Cell Press.

[12] Matsuno, H., Doi, A., Nagasaki, M. and Miyano, S. (2000). Hybrid Petri net representation of gene regulatory network. *Proc. Pacific Symposium on Biocomputing 2000*, 338–349.

[13] Matsuno, H., Murakami, R., Yamane, R., Yamasaki, N., Fujita, S., Yoshimori, H. and Miyano, S. (2003). Boundary formation by notch signaling in *Drosophila* multicellular systems: experimental observations and a gene network modeling by Genomic Object Net. *Proc. Pacific Symposium on Biocomputing 2003*, 152-163.

[14] Matsuno, H., Tanaka, Y., Aoshima, H., Doi, A., Matsui, M., Miyano, S. (2003). Biopathways Representation and Simulation on Hybrid Functional Petri Net, *In Silico Biology*, 3(3), 389 - 404, 2003.

- [15] Nagasaki, M., Doi, A., Matsuno, H. and Miyano, S. (2003). Genomic Object Net: a platform for modeling and simulating biopathways. *Applied Bioinformatics*, in press.
- [16] Proth, J.-M. (1997). Petri nets for modelling and evaluating deterministic and stochastic manufacturing systems. *Proc. 6th International Workshop on Petri Nets and Performance Models (PNPM '97)*, 2-15.
- [17] Reddy, V.N., Mavrovouniotis, M.L. and Liebman, M.N. (1993). Petri net representations in metabolic pathways. *Proc. First ISMB*, 328–336.
- [18] Reisig, W. (1985). Petri Nets. *Springer-Verlag*.
- [19] Srivastava, R., Peterson, M.S., and Bentley, W.E. (2001). Stochastic kinetic analysis of the Escherichia coli stress circuit using  $\sigma^{32}$ -targeted antisense. *Biotechnol. Bioeng.* 75(1), 120-129.
- [20] Srivastava, R., You, L., Summers, J., and Yin, J. (2002). Stochastic versus deterministic modeling of intracellular viral kinetics. *J. Theor. Biol.* 218, 309-321.
- [21] Voss, K., Heiner, M., and Koch, I. (2003). Steady state analysis of metabolic pathways using Petri nets. *In Silico Biology*, 3(3), 367-387.
- [22] Watson, J., Hopkins, N., Roberts, J., Steitz, J. and Weiner, A.M. (1987). *Molecular Biology of the Gene, Fourth edition*. The Benjamin/Cummings Publishing Company Inc.
- [23] Wheeler, G. (1999). The modelling and analysis of IEEE802.6's configuration. *Application of Petri Nets to Communication Networks (Lecture Notes in Computer Science 1605, Billington, J., Diaz, M., and Rozenberg, G. (ed.))*, 69-92.

Table 1: Transitions in Figures 7 and 9. All transitions in this figure are listed in the “Name” column. The symbol D or C in “Type” column represents the type of transition, discrete transition or continuous transition, respectively. In the “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 names of the places attached to the incoming arcs), “weight” (weight of the incoming arcs), and “type” (N, T, and I represent normal, test, and inhibitory arcs, respectively). The column “To”, which represents outgoing arcs from a transition, is divided into two sub-columns, “variable” (variable names of the places attached to the outgoing arcs) and “weight” (weight of the outgoing arcs).

Name	Type	Delay/Speed	From			To		Comment
			variable	weight	type	variable	weight	
<i>T1</i>	C	$m4/10$	$m4$	0	N	--	--	degradation rate of CAP
<i>T2</i>	C	$m14/10$	$m14$	0	N	--	--	degradation rate of mRNA repressor
<i>T3</i>	C	$m15/10$	$m15$	0	N	--	--	degradation rate of repressor
<i>T4</i>	C	$m17/10$	$m17$	0	N	--	--	degradation rate of repressor binding to DNA
<i>T5</i>	C	$m18/10$	$m18$	0	N	--	--	degradation rate of repressor not binding to DNA
<i>T6</i>	C	$m7/10$	$m7$	0	N	--	--	degradation rate of repressor binding to operator region
<i>T7</i>	C	$m19/10$	$m19$	0	N	--	--	degradation rate of <i>lacZ</i> mRNA
<i>T8</i>	C	$m20/10000$	$m20$	0	N	--	--	degradation rate of LacZ
<i>T9</i>	C	$m23/10$	$m23$	0	N	--	--	degradation rate of <i>lacY</i> mRNA
<i>T10</i>	C	$m24/10$	$m24$	0	N	--	--	degradation rate of LacY
<i>T11</i>	C	$m27/10$	$m27$	0	N	--	--	degradation rate of <i>lacA</i> mRNA
<i>T12</i>	C	$m28/10000$	$m28$	0	N	--	--	degradation rate of LacA
<i>T13</i>	C	$m29/10000$	$m29$	0	N	--	--	degradation rate of lactose outside of cell of lactose outside of cell
<i>T14</i>	C	$m9/10000$	$m9$	0	N	--	--	degradation rate of lactose
<i>T15</i>	C	$m8/2$	$m8$	0	N	--	--	degradation rate of aralactose
<i>T16</i>	C	$m30/10000$	$m30$	0	N	--	--	degradation rate of galactose
<i>T17</i>	C	$m17/10000$	$m17$	0	N	--	--	degradation rate of glucose
<i>T18</i>	C	$m10/10000$	$m10$	0	N	--	--	degradation rate of complex
<i>T19</i>	C	$m5/10000$	$m5$	0	N	--	--	degradation rate of cAMP
<i>T20</i>	C	$m11/10000$	$m11$	0	N	--	--	degradation rate of AMP
<i>T21</i>	C	$m12/10$	$m12$	0	N	--	--	degradation rate of ADP
<i>T42</i>	D	1	$m2$	1	N	--	--	CAP releasing rate
<i>T43</i>	D	1	$m3$	1	N	--	--	repressor releasing rate
<i>T45</i>	C	1	--	--	N	$m4$	--	CAP production rate
<i>T46</i>	D	1	--	--	--	$m13$	1	activation of repressor gene
<i>T57</i>	D	1.082	$m13$	1	N	$m14$	1	transcription rate of repressor
<i>T58</i>	C	$m14$	$m14$	0	N	$m15$	--	translation rate of repressor mRNA
<i>T59</i>	C	--	$m15$	--	C	$m16$	--	conformation rate of repressor
<i>T60</i>	C	$96 \times m16/100$	$m16$	0	N	$m7$	--	repressor binding rate to operator
<i>T61</i>	C	$399 \times m16/10000$	$m16$	0	N	$m17$	--	repressor binding rate to the DNA other than repressor site
<i>T62</i>	C	$m16/1000$	$m16$	0	N	$m18$	--	rate of repressor which does not bind any DNA
<i>T63</i>	D	1	$m4$	1	T	$m2$	1	binding rate of CAP to the CAP site
			$m5$	100	T			
<i>T64</i>	D	1	$m7$	1	T	$m3$	1	binding rate of repressor to the operon
			$m8$	4	I			
<i>T65</i>	D	1	$m2$	1	T	$m1$	1	logical operation of the places "CAP site" and "operator"
			$m3$	1	I			
<i>T66</i>	D	3.075	$m1$	1	T	$m19$	1	transcription rate of <i>lacZ</i>

						$m21$	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	transcription 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	transcription 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 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 allolactose from lactose inside of a cell
$T77$	D	0.5	$m3$	1	N	$m10$	1	conformation rate of repressor and allolactose
			$m8$	1	N			
			$m16$	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 allolactose from lactose outside of a cell

Table 2: 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

Table 3: Mapping between the numbers in Fig. 8 and the 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		