HYBRID PETRI NET REPRESENTATION OF GENE REGULATORY NETWORK

HIROSHI MATSUNO, ATSUSHI DOI
Faculty of Science, Yamaguchi University,
1677-1 Yoshida, Yamaguchi,753-8512, Japan

MASAO NAGASAKI
Department of Information Science, University of Tokyo,
7-3-1, Hongo, Bunkyo-ku, Tokyo, 113-0033, Japan

SATORU MIYANO
Human Genome Center, Insititute of Medical Science, University of Tokyo,
4-6-1 Shirokanedai, Minato-ku, Tokyo, 108-8639, Japan
Systems Biology Group, Kitano Symbiotic Systems Project, ERATO, JST,
6-31-15 6A, Jingu-Mae, Shibuya-ku, Tokyo 150, Japan

It is important to provide a representation method of gene regulatory networks which realizes the intuitions of biologists while keeping the universality in its computational ability. In this paper, we propose a method to exploit hybrid Petri net (HPN) for representing gene regulatory networks. The HPN is an extension of Petri nets which have been used to represent many kinds of systems including stochastic ones in the field of computer sciences and engineerings. Since the HPN has continuous and discrete elements, it can easily handle biological factors such as protein and mRNA concentrations. We demonstrate that, by using HPNs, it is possible to translate biological facts into HPNs in a natural manner. It should be also emphasized that a hierarchical approach is taken for our construction of the genetic switch mechanism of λ phage which is realized by using HPNs. This hierarchical approach with HPNs makes easier the arrangement of the components in the gene regulatory network based on the biological facts and provides us a prospective view of the network. We also show some computational results of the protein dynamics of the λ phage mechanism that is simulated and observed by implementing the HPN on a currently available tool.

1 Introduction

Some kinds of models have been studied to express gene regulatory networks or genetic networks such as electrical circuits[1], Boolean networks[2,3,4], differential equations[5,6,7,8] and stochastic Petri net [9,10]. Above all, it is widely accepted that the approach with differential equations plays a central role in modeling gene regulatory networks.

McAdams and Shapiro proposed a hybrid modeling approach that integrates conventional biochemical kinetic modeling within the framework of a circuit simulation [1]. Chen *et al.* [7] presented a number of linear differential

equation models, including both mRNA and protein levels. They showed theoretically how to solve the problem of estimating the parameters on the models using linear algebra and Fourier transforms. Novak *et al.* [6] constructed a mathematical model of fission yeast growth and division that encompasses all three crucial checkpoint controls by using a set of non-linear differential equations. A similar approach is also taken for modeling M-phase control in *Xenopus* oocyte extracts[5]. Kyoda and Kitano[8] focused on the formation of the expression patterns of some number of genes which are involved in the development of the third instar *Drosophila* leg disc. They summarized the regulatory relations of these genes as a table, represented the concentrations of proteins by differential equations, and presented a result asserting that P-D axis can be formed by a set of genes with different activation thresholds.

Now, we turn our attention to Petri net. The most attractive feature of Petri nets is that basic aspects of concurrent systems are captured conceptually as well as mathematically. A large amount of investigations on Petri nets have been compiled in the literature, and various applications have chosen Petri nets as their control models due to the intuitively understadable graphical notation of Petri nets[11].

Goss and Peccoud[9,10] introduced an approach to modeling stochastic systems in molecular biology, using stochastic Petri nets (SPNs). The approach was illustrated with examples of models of genetic and biochemical phenomena by using an existing software[9]. They also succeeded in analyzing the stabilizing effect of the protein Rom on the genetic network controlling $COLE1$ plasmid replication by using the SPN[10].

Differential equations are widely accepted to express gene regulations, but the drawback of this approach is the hardness to observe the regulation mechanism intuitively in the biological sense. On the other hand, ordinary Petri net models including SPN cannot be substituted for the differential equations in these approaches. The reason is that representing a continuous value such as the concentration of mRNA or protein is an essential factor in expressing gene regulation, but ordinary Petri net models do not have such functions. These observations led us to introduce the hybrid Petri net[12] for expressing gene regulations.

Hybrid Petri net (HPN) is an extension of Petri net that allows to handle continuous factors. The HPN enables us to express explicitly the relationship between continuous values and discrete values while keeping the characteristics of ordinary Petri nets soundly. Of course, stochastic factors, which are seemed essential for representing biological systems, can be included properly in HPNs. Moreover, the concept of HPN is well applicable to integrating conventional biochemical kinetic modeling with models of control and delay mechanisms in

gene regulatory networks. In this paper, we firstly give a method to represent gene regulatory networks by using HPNs. We then apply this method for simulating the well known genetic switch mechanism of λ phage to choose between lysis and lysogeny [13,14]. A hirarchical representation of gene regulatory networks is also realized with HPN in this paper. This hirarchical representation machanism makes it possible to obtain a prospecive view over gene regulatory networks. This will be demonstrated in Sections 3 and 4. We implemented our HPN for the genetic switch mechanism of λ phage with a tool called Visual Object Net++ [15] and the simulation results show that our approach with HPN is quite promising for further development.

2 Operon Model by Hybrid Petri Net

We assume that the reader is familiar with Petri nets [11]. In this section, we give a brief look at the definition of a hybrid Petri net together with the teminology that will be empoyed for representing the gene regulatory network of the genetic switch mechanism of λ phage. This section concentrates on the HPN representation of the operon model with which we describe the basic idea to introduce HPN for representing gene regulatory networks.

Definition 1. We denote a *hybrid PN* as $Q = (P, T, h, \mathbf{Pre}, \mathbf{Post}, M_0)$, where $P = \{P_1, P_2, ..., P_n\}(n \geq 1)$ and $T = \{T_1, T_2, ..., T_m\}(m \geq 1)$ are the sets of *places* and *transitions*, respectively;
$h : P \cup T \rightarrow \{D, C\}$ indicates for every place or transition whether it is a discrete or continuous one. A non-negative integer called the *number of token* is always associated with a discrete place ($h(P_i) = D$), and a non-negative real numbers called the *mark* is always associated with a continuous place ($h(P_i) = C$);
$\mathbf{Pre}(P_i, T_j)$ ($\mathbf{Post}(P_i, T_j)$) is a function that defines arc from a place P_i (a transition T_j) to a transition T_j (a place P_i), where the arc has a weight of non-negative integer (non-negative real number) if $h(P_i) = D$ ($h(P_i) = C$). $\mathbf{Pre}$ and $\mathbf{Post}$ functions must meet the following criterion: if P_i and T_j are a place and a transition such that P_i is discrete and T_j is continuous then $\mathbf{Pre}(P_i, T_j)=\mathbf{Post}(P_i, T_j)$ must be verified;
M_0 is a mapping from the set of places to the set of non-negative integers or the set of non-negative real numbers called the *initial marking*.

We assign a variable d_{T_j} called the *delay time of T_j* to each discrete transition T_j ($h(T_j) = D$) and assign a variable v_{T_j} called the *speed of T_j* to each continuous transition T_j ($h(T_j) = C$). In this definition of HPN, we have excluded the probabilistic feature of the system but it can be easily introduced.

With this definition of HPN, the mechanism related to transcription can be described in a natural way. Figure 1 is a hybrid Petri net which describes the transcription on an operon containing two genes, gene1 and gene2, and the translation on each of the two genes. S_1, F_1, and S_2 are discrete places, and R_1, P_1, R_2, and P_2 are continuous places. T_{R1}, T_{12}, and T_{R2} are discrete transitions, and T_{P1}, D_{R1}, D_{P1}, T_{P2}, D_{R2}, and D_{P2} are continuous transitions.

In the initial marking, the discrete place S_1 has a token (this reflects that RNA polymerase binds to the promoter of the operon), whereas the marks of other places are zero. The delay time $d_{T_{R1}}$ associated with the discrete transition T_{R1} reflects the time needed for the transcription of RNA polymerase of the gene1. (The transition T_{R1} can be fired after time $d_{T_{R1}}$ from the time when the place S_1 gets a token.) Whenever the transcription of gene1 is finished, the mark of continuous place R_1, representing the concentration of mRNA of gene1, is increased by $\mathbf{Post}(R_1, T_{R1})$. The degradation rate of concentrations of mRNA of gene1 is given by $\mathbf{Pre}(R_1, D_{R1}) \cdot v_{D_{R1}}$.

The speed $v_{T_{P1}}$ of the continuous transition T_{P1} reflects the speed of translation of gene1. (The transitions T_{P1} and D_{R1} can be fired with the speeds $v_{T_{R1}}$ and $v_{D_{R1}}$, respectively, when a mark of the place R_1 is greater than zero.) The place R_1 is simultaneously an input and an output of the transition T_{P1}, because it is required for translation but should not be consumed. The increasing rate of the concentration of protein of the gene1 (the marking of the place P_1) is given by $\mathbf{Post}(P_1, T_{P1}) \cdot v_{T_{P1}}$. The degradation rate of concentration of protein of gene1 is given by $\mathbf{Pre}(P_1, D_{P1}) \cdot v_{D_{P1}}$. The delay time $d_{T_{12}}$ of the transition T_{12} represents the time needed for RNA polymerase moving between the end of gene1 and the beginning of gene2. At the moment when the place S_2 gets a token, RNA polymerase begins the transcription of gene2. Further arguments on gene2 is omitted, since it is similar to the case of gene1 described above.

3 Growth Pathway Control of λ Phage

This section is devoted to a further HPN description of the regulatory network related to the growth pathway control by arranging discrete and continuous places and transitions with appropriate parameter adjustings. The description is given as a natural translation of the biological facts into HPN terms.

3.1 Effect of Concentrations of Two Proteins, CI and Cro

Although, for historical reasons, the cI gene product has been called 'repressor', we use the term 'CI protein' instead of this throughout this paper. It is well
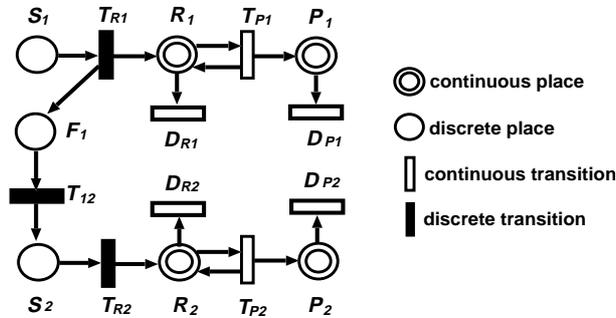
Figure 1: Two-genes operon model

known that the two regulatory proteins CI and Cro play impotant roles for deciding the lysis and lysogeny growth pathway and ultraviolet irradiation of a lysogen induces lytic growth [13,14]. The genes that encode CI and Cro are adjacent on the λ phage chromosome, and each of these genes has its own promoter, P_{RM} of the gene cI and P_R of the gene cro. The right operator of λ phage O_R consists of three adjacent sites O_{R1}, O_{R2}, and O_{R3}. CI and Cro bind to these sites to regulate the activities of the two promoters.

Table 1 summarizes the relationships among concentrations of proteins CI and Cro, binding situations of these proteins on O_{R1}, O_{R2}, and O_{R3} sites, states (ON or OFF) of the promoters P_{RM} and P_R, and the presence of ultraviolet irradiation (UV).

At the initial stage, the concentrations of both of CI and Cro are low (blanks at the column "concentration" in Table 1), and neither CI nor Cro can bind to three sites. This situation leads the promoters P_{RM} to turn off and P_R to turn on ($\odot$ and $\bigcirc$ at the column "promoter" mean ON and OFF of the promoter, respectively.)

If the concentration of CI increases to some level(+), CI ($\Diamond$) binds to two sites O_{R2} and O_{R1}. As a result, the two promoter switches are going to be reversed, that is, P_{RM} is ON and P_R is OFF. At higher concentration of CI (++), O_{R3}, as well as O_{R1} and O_{R2}, are filled. This binding of CI to O_{R3} truns off P_{RM}.

The states of the promoters P_{RM} and P_R do not change from initial stage under the condition that Cro ($\triangle$) binds to O_{R3} until the concentration of Cro is up to some level (+). However, the concentration of Cro is in exess (++), all sites are occupied by Cro. Then, both of P_{RM} and P_R are turned off.

The row in Table 1 marked '@' implies that ultraviolet light is irradiated

Table 1: Operator O_R and Promoters P_{RM} and P_R

	concentration		sites of O_R			promoter	
UV	CI	Cro	O_{R3}	O_{R2}	O_{R1}	P_{RM}	P_R
						○	⊙
	+			◇	◇	⊙	○
	++		◇	◇	◇	○	○
		+	△			○	⊙
		++	△	△	△	○	○
@	*	*	*	*	*	○	⊙

for a period of time. Ultraviolet irradiation of lysogens inactivates CI protein. Thereafter, Cro protein is begun to synthesize. It follows that, independently of the concentration and the binding situation of proteins CI and Cro (*), the promoter P_R is going to be ON.

Figure 2 shows an HPN representation of Table 1. Two places are continuous, and others are discrete. All transitions are discrete. The weight $\mathbf{Pre}(CI,A_{CI})=\mathbf{Post}(CI,A_{CI})$ ($\mathbf{Pre}(Cro,A_{Cro})=\mathbf{Post}(Cro,A_{Cro})$) indicates a threshold level of CI (Cro) concentration that CI (Cro) binds to the sites O_{R2} and O_{R1} (the site O_{R3}). The weight $\mathbf{Pre}(CI,B_{CI})=\mathbf{Post}(CI,B_{CI})$ ($\mathbf{Pre}(Cro, B_{Cro})=\mathbf{Post}(Cro,B_{Cro})$) indicates a threshold level of CI (Cro) concentration that CI (Cro) binds to the site O_{R3} (the sites O_{R2} and O_{R1}).

From Table 1, we can easily see that it is not needed to distinguish CI ($\diamondsuit$) and Cro ($\triangle$) for getting the required outputs of P_{RM} and P_R. Parameters of transitions D_{OR3}, D_{OR2}, and D_{OR1} represent the degradation rate of CI and Cro at the sites O_{R3}, O_{R2}, and O_{R1}, respectively.
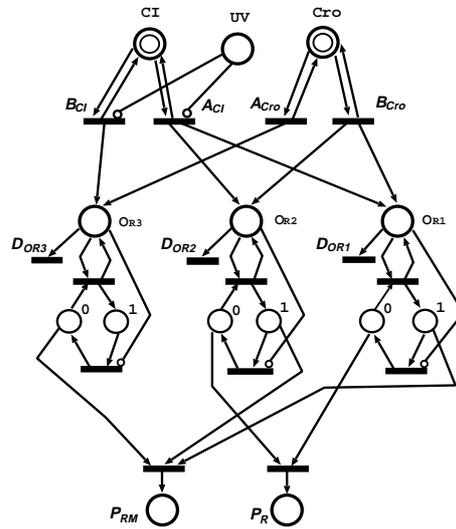
It is easy to verify that dynamics of Figure 2 exactly corresponds to Table 1.

3.2 Feedback Mechanism of CI and Cro

The factor that determines which pathway of lysogen and lysis is selected is the concentration of protein CII; that is, if CII is highly active then the infecting phage lysogenizes; otherwise it grows lytically. It is known that the activity of CII is determined by environmental factors. Furthermore, two proteins CI and Cro are also the keys to determine the pathway of lysogeny or lysis.

Figure 3 describes this mechanism by an HPN. In the square surrounded by dotted line in the figure, the operator O_R shown in Figure 2 should be inserted.

By referring to the previous section, we can easily see that each of non-

Figure 2: Operator O_R

labeled continuous transitions has the parameter of either mRNA degradation, protein degradation, or protein synthesis rate.

The gene cI can be transcribed from either of two promoters, one is P_{RM} activated by CI, the other is P_{RE} activated by CII. If the concentration of CII is high (the threshold is given by $\mathbf{Pre}(\text{CII}, A_{CII})$) and the promoter P_{RE} is going to be ON, then the concentration of CI keeps growing during the promoter P_{RM} is ON. In addition, the transcript initiated at P_{RE}, which includes 'anti-sense' cro sequences, hybridizes with cro mRNA and prevents its translation. The parameter of discrete transition labeled D_{cro} reflects the rate of this hybridization. At very high concentration of CI, P_{RM} switches to OFF. It follows that its own overproduction is prevented as is seen in Table 1. Synthesis of the CI is therefore self-regulated both positively and negatively.

During the concentration of Cro is up to a level, the promoter P_R keeps ON, and Cro protein is continuously produced until it reaches overproduction level. Note that these levels are given as weights of arcs in the way as we showed at the previous section.

The place labeled by C_{RO_E} is described for the convenience of the discussion in the next section. The place indicates the termination of transcription of the gene cro.
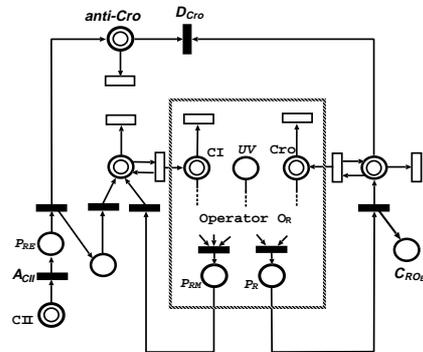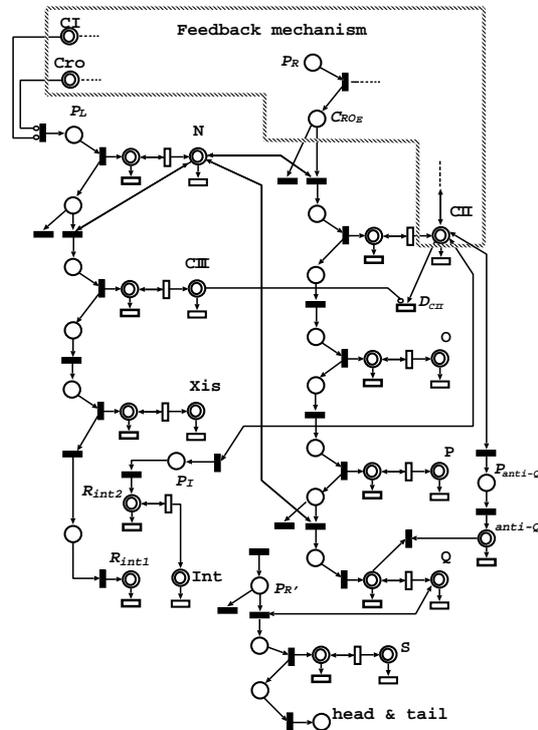
Figure 3: Feedback mechanism of CI and Cro

4   HPN Representation of Early Stage Gene Expressions of λ Phage

Figure 4 shows the gene regulatory network that determines the growth path of λ phage by the HPN. In the square surrounded by dotted line in the figure, the feedback mechanism of CI and Cro shown in Figure 3 should be inserted.

P_L operon has the genes N, $cIII$, xis, and int. P_R operon has the genes cro, cII, O, P, and Q. Note that these operons are described in the figure in the same manner as we gave in Section 2. $P_{R'}$ operon has the gene S and downstream genes that encode cell lysis proteins and head and tail coat proteins.

N protein regulates early gene expression by acting at three terminators: one between itself and the neighbor gene $cIII$, one between genes cro and cII, and one between genes P and Q. CII turns on cI and int. It encourages RNA polymerase to bind and begin transcription at two promoters that would remain silent: P_{RE} and P_I. CIII protein helps to establish lysogeny, that is, its role is to protect CII from degradation. Int protein helps to integrate the chromosome of λ phage into the host chromosome. In the case of reverse reaction, excision, the protein Xis is needed in addition with Int. Protein O and P proteins are required for DNA replication. Q protein turns on the late genes–those for lysis and for production of heads and tails. It anti-terminates specifically a small RNA begun at a promoter $P_{R'}$, located just to the neighbor of Q.

The marks of continuous places labeled with the names of proteins represent the concentrations of these proteins. The bi-direction arcs between the place labeled N and three discrete transitions indicate that N anti-terminates

Figure 4: Early stage gene expressions of λ phage

the terminators to allow transcriptions of genes followed by the three transitions. The bi-direction arc between the place labeled Q and the discrete transition has the same meaning as above. These bi-direction arcs have the weights representing the thresholds of concentration enough for the anti-terminations, although these are not represented explicitly in the figure.

An observation concerning to the decision of lysogeny or lysis pathway is given in [14], that is, not only CII protein effects the decision but also CIII protein does. In addition to the natural protein degradation rate, we assume that CII protein makes extra degradation by the rate of speed of the transition D_{CII}. If the concentration of CIII exceeds the level given by **Pre**(CIII,D_{CII}), the arc between CIII and D_{CII} inhibits the extra degradation of CIII protein.

Under the conditions favoring CII protein activity, the concentrations of CI and Int grow rapidly. The *int* gene is transcribed from the promoter P_L

as well as from the promoter P_I. However, the mRNA of *int* initiated at P_L is degraded by cellular nucleases, whereas mRNA initiated at P_I is stable and can be translated into the protein Int. Note that in the case of induction of a lysogen, it is known that the transcription from P_L is needed for making Xis and Int, but details are omitted here. The places labeled by R_{int1} and R_{int2} represent the concentrations of *int* mRNAs initiated at P_L and P_I, respectively. We do not describe the arcs and transition for translation between R_{int1} and Int, because we can see that these are not needed from the argument above.

CII inhibits expression of the late genes by stimulating a promoter called P_{anti-Q}. That promoter directs backwards transcription of the Q gene. The anti-sense Q RNA hybridizes with and prevents its translation. We can see that this function is reflected in the Figure 4.

At the first, RNA polymerases bind to three promoters P_L, P_R, and $P_{R'}$ and begin transcription. It follows that, at the initial marking, only each of the places corresponding to the three places has a token. The parameters of transitions are selected to appropriate values so that these are consistent with the biological phenomena described in [13,14].

We have implemented the gene regulatory network presented in this paper by using Petri-Net-CAD/CAE-Tool called Visual Object Net++ which works on Windows95/NT [15]. Of course, the tool supports mixed continuous and discrete event Petri nets. Since the tool does not support hierarchical concepts, we made simulations on the Petri net integrating Figures 2, 3, and 4.

The dynamics of protein concentrations which are obtained by simulating the hybrid Petri net described in this paper are shown in Figure 5. These results are consistent with the biological facts well [13,14]. The files which we made can be downloaded from the website [16]. It is known that the choice of either lysogeny or lysis is determined by the combination of CII and CIII proteins. Figure 5 shows the cases that the rate of CII concentrations are different, whereas the rate of CIII concentrations the are same. The left and half parts of the figure show the cases which establish the lysogeny and lysis paths, respectively.

If the concentration of CII keeps high rate, it reaches the threshold levels enough to stimulate the promoters P_{RE}, P_I,and P_{anti-Q}. Then, the concentrations of CI and Int are going to grow, and the transcription of 'anti-sense' Q RNA is begun. By the feedback mechanism stated in Section 3, the concentration of CI keeps a level enough for changing to the lysis growth by ultraviolet irradiation. We can also observe that the activities of Q protein is repressed by the anti-sense Q RNA.

In contrast with this, if the concentration of CII is low, the promoters P_{RE} and P_{RM} are never turned on, but the promoter P_R is turned on. As a
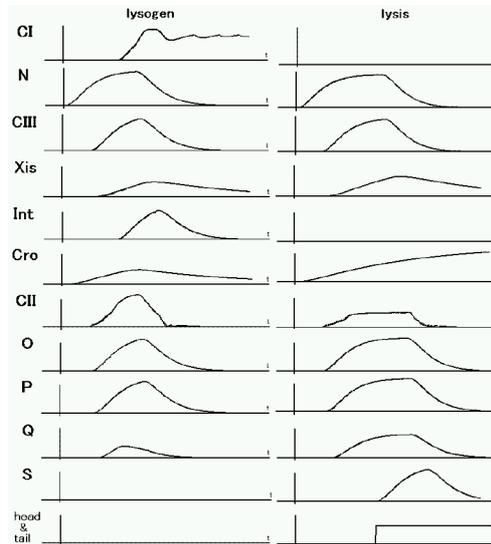
Figure 5: Dynamics of protein concentrations.

result, the concentration of Cro protein keeps increasing. The Q protein anti-terminates the terminator just located at the neighbor of the P_R, then the S protein is synthesized and the production of heads and tails is begun.

5 Conclusion

We believe that the method introduced in this paper can provide a powerful test of the hypothesis which biologists are going to set up, because the graphical representation of proteins and mRNAs in the HPN is similar to standard representations of them in biochemistry. Of course, the HPN can handle probabilistic factors in biological phenomena. Moreover, we demonstrate how gene regulatory networks are represented hierarchically. In addition to the merit of arranging the view of the gene regulatory networks, hierarchical representations enable us to place a black box at the location where the detailed mechanism is unknown but the relation of input and output is known as a function.

McAdams and Shapiro[1] proposed a hybrid modeling approach that integrates conventional biochemical kinetic modeling within the framework of a electrical circuit simulation. However, basically, their model consists of two different kinds of parts, circuit diagrams and differential equations. In this pa-

per, we show that HPN can integrate such different kinds of parts and perform the direct simulation on the HPN representing a gene regulatory network by using a currently available tool. Simulation results of another organisms such as $Xenopus$[5] and fission yeast[6] will be reported elsewhere soon.

1. H. H. McAdams and L. Shapiro, Circuit simulation of genetic networks, *Science* **269**, .650-656, 1995.
2. S. Liang, S. Fuhrman, and R. Somogyi, REVEAL, a general reverse engineering algorithm for inference of genetic network architectures, *Pacific Symposium on Biocomputing* **3**, 18-29, 1998.
3. T. Akutsu, S. Kuhara, O. Maruyama, and S. Miyano, Identification of gene regulatory networks by strategic gene disruptions and gene overexpressions, *Proc. 9th ACM-SIAM Symp. Discrete Algorithms*, 695-702, 1998.
4. T. Akutsu, S. Miyano, and S. Kuhara, Identification of genetic networks from a small number of gene expression patterns under the Boolean network model, *Pacific Symposium on Biocomputing'99,* 17-28, 1999.
5. G. Marlovits, C.J. Tyson, B. Novak, and J.J. Tyson, Modeling M-phase control in *Xenopus* oocyte extracts: the surveillance mechanism for unreplicated DNA, *Biophysical Chemistry* **72**, 169-184, 1998.
6. B. Novak, A. Csikasz-Nagy, B. Gyorffy, K. Chen, and J. J. Tyson, Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metaphase/anaphase transitions, *Biophysical Chemistry* **72**, 185-200, 1998.
7. T. Chen, H. L. He, and G. M. Church, Modeling gene expression with differential equations, *Pacific Symposium on Biocomputing'99,* 29-40, 1999.
8. K. Kyoda and H. Kitano, Simulation of genetic interaction for *Drosophila* leg formation, *Pacific Symposium on Biocomputing'99,* 77-89, 1999.
9. P. J. E. Goss and J. Peccoud, Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets, *Proc. Natl. Acad. Sci. USA,* **95**, 6750-6755, 1998.
10. P. J. E. Goss and J. Peccoud, Analysis of the stabilizing effect of Rom on the genetic network controlling *ColE1* plasmid replication, *Pacific Symposium on Biocomputing'99,* 65-76, 1999.
11. W. Reisig and G Rozenberg (Eds.), Lectures on Petri nets I:Basic models, *Lecture notes in Computer Science* **1491**, 1998.
12. H. Alla and R. David, Continuous and hybrid petri nets, *Journal of Circuits, Systems, and Computers,* **8(1)**, 159-188, 1998.
13. J.D. Watson, N. H. Hopkins, J. W. Roberts, J. A. Steitz, and A. M. Weiner, *Molecular biology of the gene, fourth edition*, The Benjamin/Cummings Publishing Company, Inc., 1988.
14. M. Ptashne, *A genetic switch phage λ and higher organisms, second edition*, Cell Press & Blackwell Science, 1992.
15. *http : //www.systemtechnik.tu-ilmenau.de/˜drath/visual_E.htm*
16. *http : //genome.ib.sci.yamaguchi-u.ac.jp/˜atsushi/phage/sim.html*