XML Documentation of Biopathways and Their Simulations in Genomic Object Net

Hiroshi Matsuno[1]
matsuno@sci.yamaguchi-u.ac.jp

Atsushi Doi[1]
atsushi@ib.sci.yamaguchi-u.ac.jp

Yuichi Hirata[2]
hirata@ims.u-tokyo.ac.jp

Satoru Miyano[2]
miyano@ims.u-tokyo.ac.jp

[1] Faculty of Science, Yamaguchi University, 1677-1 Yoshida, Yamaguchi 753-8512, Japan
[2] Human Genome Center, Institute of Medical Science, University of Tokyo, 4-6-1 Shirokane-dai, Minato-ku, Tokyo 108-8639, Japan

Abstract

Genomic Object Net is a software tool for modeling and simulating biopathways which employs the notion of hybrid functional net as its basic architechture. This paper shows how to integrate this basic architecture with XML documents for biopathway representations, simulations, and visualizations for creating a tailor-made simulation environment.

Keywords: hybrid functional Petri net, XML, biopathways, Genomic Object Net, visualization

1 Introduction

Biosimulation systems have received considerable attentions [7, 3, 13, 12, 9, 6] from a new interest of research. It aims at developing information technology with which we can easily represent and simulate complex biological systems and apply the technology to medicine and biology. As biological systems, we consider that the first stage of research should deal with biopathways which include dynamic causal interactions and processes of various biological objects such as genomic DNA, mRNA, proteins, functional proteins, molecular transactions and processes such as gene regulations, signal transduction cascades, metabolic pathways, etc.

In order for software tools to be accepted by users in biology/medicine for biopathway representation and simulation, the following matters should be resolved, at least:

(1) Remove issues that are irrelevant to biological importance.

(2) Allow users to represent biopathways intuitively and understand/manage easily the details of representation and simulation mechanism.

(3) Provide users an environment where users can visualize, interpret, evaluate the simulation results and test hypotheses.

We have developed a software tool Genomic Object Net for representing and simulating such interactions and processes. For this purpose, we introduced the hybrid functional Petri net (HFPN) model as its basic architecture for simulation mechanism. The HFPN model is defined as an extention of the hybrid Petri net model (HPN) [1]/hybrid dynamic net (HDN) [2] and we have demonstrated [6, 5] that with this model we can represent intuitively and simulate natually typical biopathways related to gene regulation (λ phage switching mechanism, circadian rhythm of *Drosophila*), metabolic pathway (lac operon and glycolytic pathway of *E. coli*), and signal transduction (apoptosis induced by Fas ligand), which cover the basic aspects in biopathways. Genomic Object Net has HFPN-based

tools *Biopathway Describing Tool* and *Genomic Object Net Assember* which work together for realizing (1) and (2).

It is natural to use XML documents for describing HFPN-modeled biopathways, visualizing biopathways, thier simulations, and related information such as database links. The purpose of this paper is to show how these issues are integrated in Genomic Object Net and to see that with Genomic Object Net, users can construct a tailor-made simulation environment for their specific purpose. As an example of such biopathway XML document, we will exhibit a full series of animated simulations of the λ phage gene regulatory network [10] that is represented and simulated with Genomic Object Net. The software and representation/simulation examples are freely available from `http://www.GenomicObject.Net/`.

2 Hybrid Functional Petri Net for Representing Biopathways

2.1 Design Example of λ Phage Genetic Switch Mechanism on Hybrid Petri Net

In [6], it is demonstrated that HPN has the excellent ability to represent biological processes such as transcription, translation, and protein binding through the description on λ phage genetic switch mechanism and the simulation of protein concentration produced by the mechanism. Figure 1 is described for the intuitive understanding of the readers.

Continuous places (doubled circles) contain real numbers which represent concentrations of mRNA and proteins. Discrete places (single circles) are corresponded to the promoter or operator sites for expressing the binding situation of transcription factors, and also used for representing whether RNA polymerase is located between genes.

Continuous transitions (white rectangles) are used for expressing the production/degradation speed of mRNAs or proteins. Discrete transitions (black rectangles) are used for expressing the required time for RNA polymerase which moves between two genes, and also used for representing CI (Cro) binding condition on the three operator sites depending on concentration of itself.

Inhibit arc (arc with small circle) is used for representing the biological process which relates to repression functions of gene products.

2.2 Hybrid Functional Petri Net

HPN is quite promising for expressing biopathways, however, technical expressions of HPN still have some difficulties for representing some kinds of biological processes, and these difficulties can not be resolved even if we apply HDN, which can make more descriptive expressions than HPN, to express such biological processes.

By resolving these difficulties in a systematic way, we introduced a notion of *hybrid functional Petri net* (HFPN) [5] for biopathway modeling by extending the definition of transition. Intuitively speaking, a transition in the HFPN model can control the speed of flow in each arc by a program while the HPN and HDN allow the same flow speed in each arc connected to the transition.

For explaining the usefulness of HFPN briefly, we present a simple example in the following. For biopathway simulations, even the HDN model, which is the most general model, is not convenient enough. Consider a reaction which decomposes dimers to monomers (Figure 2 (a)). This reaction can be represented as shown in Figure 2 (b) by using a test arc and a transition for amplification (note that the amounts consumed and produced in places by continuous transition firing is the same by definition while the amount of monomers is twice as large as that of dimers). But it is neither intuitive nor natural at all. On the other hand, in the HFPN, it can be represented as Figure 2 (c).

The effectiveness of representing ability of HFPN on descriptions of biological systems has been already confirmed through examples presented in [5].

Genomic Object Net Assembler implemented the HFPN as the basic representation and simulation architecture together with a graphical visualization tool [5]. With this tool, an interesting software

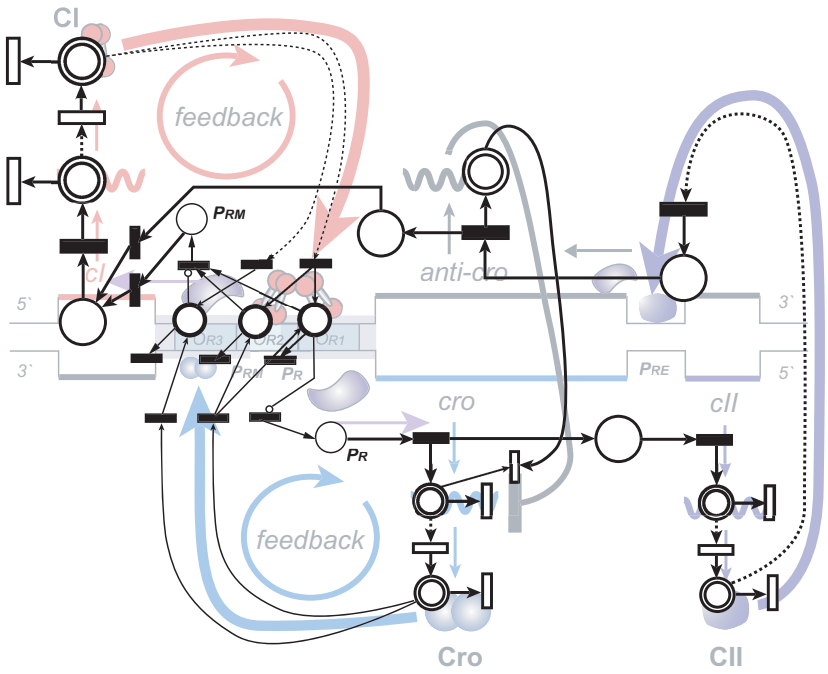


Figure 1: HDN description on the example of λ phage genetic switch feedback mechanism: Transcription of the genes *cro*, *cII* and genes followed by *cII* gene from the promoter P_R begin, when neither CI protein nor Cro protein does not bind to the operator sites O_R3, O_R2, and O_R1. Condition of *E.coli* effects to concentration of CII protein. If the concentration of CII protein is low, the transcription from P_R continues and keeps the concentration of Cro protein at some level by the feedback control of the Cro protein itself. On the other hand, if the concentration of CII protein is high, the CII protein binds to the promoter P_{RE} as a positive transcription factor, then transcription from P_{RE} begins. Then, anti-sense RNA of the gene *cro* is produced, which helps to degrade the concentration of Cro protein more rapidly. Transcription of *cI* gene is followed and concentration of CI protein keeps at some level by the feedback control of the CI protein itself.

tool such as E-Cell [13] can be realized as a subset of Genomic Object Net.

3 XML Integration of Biopathway Representations, Simulations, and Visualizations Genomic Object Net

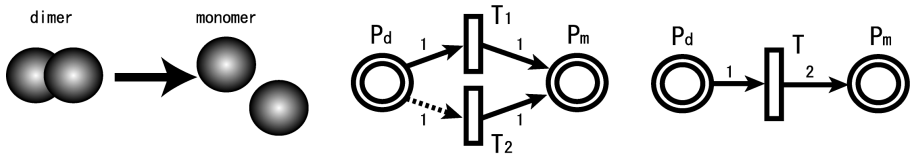
Our strategy of biopathway representation and simulation mainly consists of the following two items:

1. HFPN for modeling, representing, and simulating biopathways.
2. XML documents for description and visulalization of biopathways and their simulations.

The purpose of this section is to show how these two items are integrated in Genomic Object Net. Figure 3 gives the conceptual relationship between issues involved in Genomic Object Net.

3.1 Biopathway Describing Tool and Genomic Object Net Assembler

"Biopathway Describing Tool" in Figure 3 is a tool which helps to describe a biopathway by following the HFPN design principle in Section 2.2. It aims at releasing users from the descriptive and mathematical knowledge such as programming languages and differential equations, which are inevitable



(a) Reaction decomposing dimers to monomers (b) hybrid Petri net (c) hybrid functional Petri net

Figure 2: (a) is a reaction model decomposing dimers to monomers. (b) and (c) are the HDN and HFPN representations of (a), respectively. Continuous places P_d and P_m in (b) and (c) contain the numbers of dimers and monomers, respectively. The HFPN of (c) realizes the reaction naturally, in which two monomers are created by dividing one monomer, by assigning the weights 1 to the arc $P_d \rightarrow T$ and 2 to the arc $T \rightarrow P_m$. That is, the amount flows through the arc $T \rightarrow P_m$ at the double speed of the one through the arc $P_d \rightarrow T$. However, since such function is not allowed to HDN, the description of HDN (b) becomes more complicated than that of HFPN (c).

issues in most simulation systems [7, 13]. This tool realizes a platform on which biopathways can be represented and simulated in harmony with biologically intuitive understanding of the biopathways.

"Genomic Object Net Assembler" shall cooperate with "Biopathway Describing Tool" in the following ways:

1. Design of an HFPN which models a biopathway.
2. Production of a file called *biopathway representation source file* that describes the HFPN as an XML document.
3. Production of a CSV format file which contains a time-series of the numerical values representing the behavior of places in the HFPN in the simulation.

These two tools are combined into a system for editing and simulating biopathways in Genomic Object Net.

3.2 HFPN desrciption in XML

The HFPN desrciption in XML format can be easily and naturally realized as in the following example. We skip the details about this format.

```
<?xml version="1.0" encoding="UTF-8"?>
<!DOCTYPE hybridFunctionalPetriNet SYSTEM "SampleHFPNet.dtd">
<HFPN>
  <place id="P1" type="continuos" variableName="m1"/>
  <place id="P2" type="continuos" variableName="m2"/>
  <transition id="T1" speedFunction="m1/2.5" type="continuos"/>
  <arc from="P1" to="T1" type="normal" weight="1"/>
  <arc from="T1" to="P2" type="normal" weight="m1/2"/>
</HFPN>
```

3.3 Biopathway XML Document

If user need to interprete/evaluate the biopathway behavior in a purpose-specific way, user can create an XML document for the purpose (we call this file *biopathway XML document*) by using a biopathway representation source file (mechanism), CSV files (simulation results), various other data files

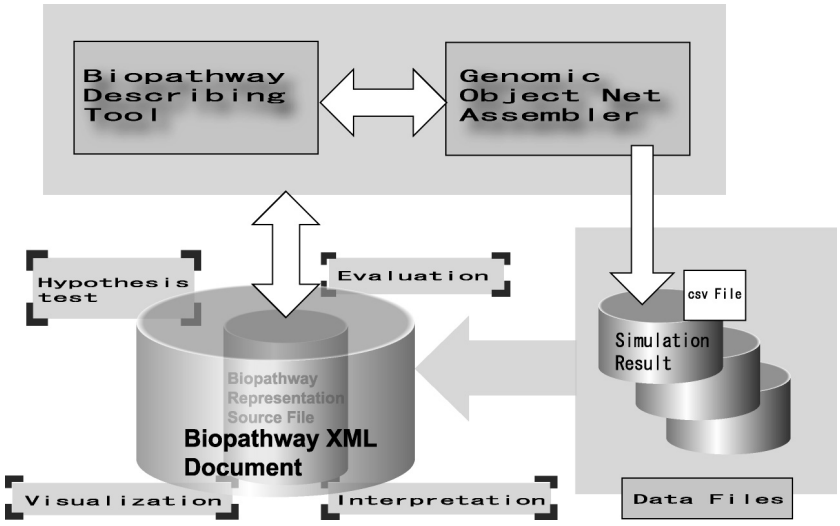


Figure 3: "Biopathway Describing Tool" is a tool which describes the aimed biopathway. "Genomic Object Net Assembler" cooperates with Biopathway Describing Tool in constructing HFPN on which genomic objects work and in producing the biopathway representation source file as an XML document. Genomic Object Net Assembler produces the CSV format files (simulation results) each of which contains the numerical values representing behavior of genomic object to change with time such as the concentration of protein. A biopathway XML document is the base and central file of the whole system. All information given to/produced by the system are integrated in the file so that user can make visualization, evaluation, interpretation, or hypothesis test of their aimed biopathway on the system, effectively and smoothly.

including XML and DTD files (visualization, animation, database links, etc.), pictures depicted with general purpose drawing tools (encouragement of intuitive understanding), etc. Namely, a biopathway XML document is a file where all information regarding the biopathway of interest is systematically integrated. With this file, user can make visualization/animation, evaluation, interpretation, or hypothesis test of the HFPN-represented biopathway effectively and smoothly in a tailor-made way. An example is shown in the next section.

4 XML Biopathway Document for Visualization

This section demonstrates, by using the λ phage gene regulatory network as an example, how a biopathway XML document is realized in Genomic Object Net. Here we assume that the purpose of constructing this XML document is to show some biological processes in an animated way so that their simulated behaviors will realize the descriptions in the biological literature [10]. More precisely, we intended to create a biopathway XML document which realizes a screen shot of the animation of the λ phage genetic switch mechanism behavior as shown in Figure 4. In the following, we shall sketch some parts of the XML document to see how it is created.

4.1 Data Files

Firstly, CSV files containing simulation time-series data are created from "Genomic Object Net Assember" by running the HFPN which represents the λ phage gene regulatory network whose primary version was presented in Matsuno *et al.* [6]. When creating a CSV file, we specify the places of interest for visualization/evaluation. Secondly, by using a general-purpose drawing tool, appropriate pictures

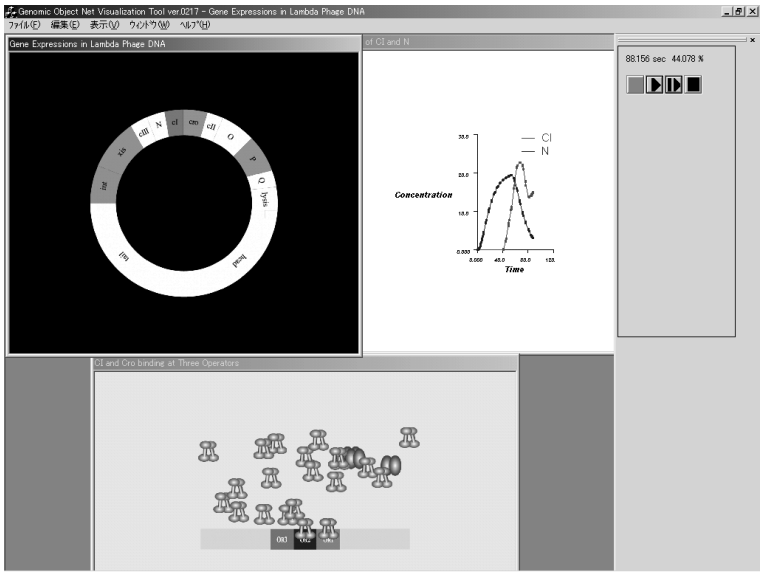


Figure 4: Screen shot of a visualization example of a biopathway XML document: The upper left window shows the λ phage chromosome in which intensities or colors of genes are changed according to the protein concentration levels produced by these genes. The upper right window shows concentration behaviors of the proteins N and CI. The bottom window shows binding situation of CI and Cro proteins to the three operator sites O_R1, O_R2, and O_R3. Animations in these windows are driven by values in a CSV file containing simulation results.

which would help visual understanding are created as PNG files for the specified places in the HFPN and the background pictures. Then the XML document will use these CSV files and picture files for driving and visualizing the animation parts.

4.2 XML Tags for Visualization

In the following, by focusing on the behaviors of CI and Cro proteins at the bottom of Figure 4, we will explain some selected tags of XML document which we have developed for creating this XML document. The whole XML document with its manual can be downloaded from `http://www.GenomicObject.Net/`.

Speed of simulation can be controlled by two attributes of `animationClock` and `physicalClock` in the tag;

```
<simulation animationClock="0.8" physicalClock="0.2">.
```

The PNG format image files of two proteins CI and Cro and three operators O_R1, O_R2, and O_R3 in the figure are specified by the attribute `file` in the `imageFile` tag with the attribute of `id`, like;

```
<imageFile id="c1-protein-image" file="..\picture\repressor.png"/>.
```

The XML document identifies the Simulation Results File by the following tag with two attributes `file` and `id`;

```
<sequenceFile id="CI-event" file="CI.csv"/>,
```

this is an example where CI.csv is identified as the `id` name `CI-event`.

The window of animation is created by the following tag;

```
<animation title="CI and Cro binding at Three Operators"
backColor="0.9 0.9 0.9">,
```

where the window name is specified by the attribute `title` and the window background color is determined by the attribute `backColor`.

The tag `<animeStraightSequence>` is used for composing the objects in a straight, and these objects are specified by the tag `<amineStructureElement>`. The following example constructs the line of operators O_R1, O_R2, and O_R3 described in the bottom window in Figure 4,

```
<animeStraightSequence id="operators" size="200 20" >
     <animeStructureElement id="or3" image="or3-image">
     </animeStructureElement>
     <animeStructureElement id="or2" image="or2-image">
     </animeStructureElement>
     <animeStructureElement id="or1" image="or1-image">
     </animeStructureElement>
</animeStraightSequence> ,
```

where, in `<animeStraightSequence>` tag, the attribute `id` assigns a name to the composed object and the attribute `size` specifies the size of composed object. The attribute `image` in `<animeStructureElement>` tag identifies the image file of aimed object, and if this object is refereed in a tag in the same XML document, the attribute `id` should be used, otherwise, this attribute is not needed.

Two proteins CI and Cro are moved in the bottom window of Figure 4. The following XML description is the part of XML document, in which the movements of CI proteins are controlled.

```
<animeMultiObject image="c1-protein-image" size="20 20"
    lowerBound="-100 20 0" upperBound="100 100 0">
    <animeMultiVisibility event="CI-event" discretization="0.5"
       threshold="3"/>
    <animeMultiMotion event="OR1_CI-event" type="reversible"
       motionPeriod="1" reference="or1" position="0 10 0"/>
</animeMultiObject>
```

The tag `<amineMultiObject>` is used in the situation that more than one object should be displayed. The attribute `image` specifies the picture of object which was defined in the `<imageFile>` tag and the attribute `size` specifies the size of the object. Objects are displayed in the area whose range is defined by the attributes `<UpperBounds>` and `<lowerBound>` in the tag.

Visual conditions of objects are defined by the tag `<animeMultiVisibility>`. This tag with the attributes `event="CI-event"`, `discretization="0.5"`, and `threshold="3"` means that CI proteins are displayed when the value of `CI-event`, which was specified in the `<sequenceFile>` tag, exceeds the threshold 3, and a number of CI proteins are determined by the formula of these attributes and the attribute `discretization`, `(event-threshold)/discretization`.

The motion of object is controlled by the tag `<animeMultiMotion>`. This tag with the attributes,

```
event="OR1_CI-event", type="reversible", motionPeriod="1",
reference="or1"+, \verb+position="0 10 0"
```

has the meaning that CI proteins goes and returns (`"reversible"`) between the position at `"or1"` located and the position of `"0 10 0"` at intervals of `"1"` according to the value of `"OR1_CI-event"`.

We have not enhanced database links in this XML documents and another data useful for understanding because of simplicity of explanation. However, it is obvious that very rich information can be embedded in its XML biopathway file very easily.

5 Discussion

One of the essential requirements for software tools to be accepted by users in biology/medicine for biopathway representation and simulation is to remove issues which are irrelevant to biological importance. The notion of HFPN introduced in [5] satisfy this requirement, since it allows users to represent biopathways intuitively and understand/manage easily the details of representation and simulation mechanism. More precisely, using HFPN as representing method for biopathways makes it possible to hide mathematical principle and descriptive method such as differential equations and programming languages from users in biology/medicine.

We believe that the concept of hiding irrelevant issues in modeling biological phenomenon is essentially required for representing biosimulation systems. The currently available biosimulation tools such as GEPASI [7], E-Cell [13], and BioDrive [9] are not enough to satisfy this requirement since skills at modeling biopathway by differential equations and/or implementing models in a computer by some programming language are needed for handling these tools effectively.

This paper presented one more application of hiding concept stated above; XML integration of biopathway representations, simulation, and visualization.

A Document Type Description (DTD) defines the grammar of XML document by specifying allowable combinations and nesting of tag names, attribute names, and so on. The concept of XML and DTD structure is quite successful in representing biopathways, since this structure enable us to separate complicated mathematical descriptions to the DTD file from biologically essential descriptions. This concept contributes to reduce efforts of users in biology/medicine to learn programming languages which are basically computer technology, but to put it the other way around, computer scientists should make efforts to create DTD files based on the discussion with the users in biology/medicine.

6 Conclusion

Genomic Object Net is a software which can create a user-designed biopathway simulation environment by realizing the following two features:

1. The HFPN model for representing and simulating biopathways.
2. XML documents for description and visualization of HFPNs together with simulation results.

The future plan of Genomic Object Net is to construct a knowledge database of biopathways which allow computer simulation and visualization.

Acknowledgments

This work is partially supported by the Grand-in-Aid for Scientific Research on Priority Areas (C) "Genome Information Science" and Grand-in-Aid for Scientific Research (B) (No.12480080) from the Ministry of Education, Culture, Sports, Science and Technology in Japan

References

[1] Alla, H. and David, R., Continuous and hybrid Petri nets, *Journal of Circuits, Systems, and Computers*, 8(1):159–188, 1998.

[2] Drath, R., Hybrid object nets: an object oriented concept for modeling complex hybrid systems, *Proc. Hybrid Dynamical Systems, 3rd International Confernce on Automation of Mixed Processes, ADPM '98*, 437–442, 1998.

[3] Heidtke, K.R. and Schulze-Kremer, S., Design and implementation of qualitative simulation model of λ phage infection, *Bioinformatics*, 14:81–91, 1998.

[4] Hofestädt, R. and Thelen, S., Quantitative modeling of biochemical networks, *In Silico Biology*, 1(1):39–53, 1999.

[5] Matsuno, H., Doi, A., Tanaka, Y., Aoshima, H., Hirata, Y., and Miyano, S., Genomic Object Net: its basic architecture for representing and simulating biopathways, *submitted*.

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

[7] Mendes, P., GEPASI: a software for modeling the dynamics, steady states and control of biochemical and other systems, *Comput. Appl. Biosci.*, 9(5):563–571, 1993.

[8] Koch, I., Schuster, S., and Heiner, M., Simulation and analysis of metabolic networks by time-dependent Petri nets, *German Conference on Bioinformatics Poster Abstracts*, 1999. http://www.bioinfo.de/isb/gcb99/poster/koch/

[9] Kyoda, K., Muraki, M., and Kitano, H., Construction of a generalized simulator for multi-cellular organisms and its application to Smad signal transduction, *Pacific Symposium on Biocomputing 2000*, 317–329, 2000.

[10] Ptashne, M., *A Genetic Switch Phage λ and Higher Organisms*, second edition, Cell Press & Blackwell Science, 1992.

[11] Reddy, V.N., Liebman, M.N., and Mavrovouniotis, M.L., Qualitative analysis of biochemical reaction systems, *Comput. Biol. Med.*, 26:9–24, 1994.

[12] Shaff, J. and Loew, L.M., The virtual cell, *Pacific Symposium on Biocomputing 1999*, 228–239, 1999.

[13] Tomita, M., Hashimoto, K., Takahashi, K., Shimizu, T., Matsuzaki, Y., Miyoshi, F., Saito, K., Tanida, S., Yugi, K., Venter, J.C., and Hutchison, C., E-CELL: Software environment for whole cell simulation, *Bioinformatics*, 15:72–84, 1999.