Journal of Bioinformatics and Computational Biology
Vol. 2, No. 4 (2004) 619–637

Imperial College Press
www.icpress.co.uk

MODELING AND SIMULATION OF MOLECULAR BIOLOGY SYSTEMS USING PETRI NETS: MODELING GOALS OF VARIOUS APPROACHES

SIMON HARDY
Computer Engineering Department, École Polytechnique de Montréal
P.B. 6079, succ. Centre-Ville
Montréal, Québec H3C 3A7, Canada
simon.hardy@polymtl.ca

PIERRE N. ROBILLARD
pierre-n.robillard@polymtl.ca

Received 24 October 2003
Revised 10 February 2004
Accepted 29 March 2004

Petri nets are a discrete event simulation approach developed for system representation, in particular for their concurrency and synchronization properties. Various extensions to the original theory of Petri nets have been used for modeling molecular biology systems and metabolic networks. These extensions are stochastic, colored, hybrid and functional. This paper carries out an initial review of the various modeling approaches based on Petri net found in the literature, and of the biological systems that have been successfully modeled with these approaches. Moreover, the modeling goals and possibilities of qualitative analysis and system simulation of each approach are discussed.

Keywords: Biochemical pathways modeling; petri net; qualitative analysis; simulation.

1. Introduction

The completion of the human genome sequencing project, the rapid development of bioinformatics and the phenomenal accumulation of biological data have made the understanding biological processes and cellular functions a growing research interest. This post-genomic wave is called "systems biology", and with it has arisen a greater interest in the modeling and simulation of biological systems.[1] Many formalisms from the fields of biology, mathematics and the computer sciences are used to integrate, represent and analyze the vast amount of biological data.

A traditional representation uses ordinary differential equations (ODEs) to model biological systems. It is widely used and many tools are based on this

approach.[2] Appropriate for representing and simulating the kinetic equations of biochemical reactions, ODEs are mostly used to study the dynamics of a metabolic process. The software packages Gepasi[3] and E-CELL[4] have been developed to support modeling with this analytical representation. Boolean logic and state machines are also used for modeling biological systems. Even though boolean models are simple representations, they can be used to express characteristics of biological phenomena.[5] For example, a logic model of the *Endo16* gene of *Strongylocentrotus purpuratus* predicted an internal *cis*-regulatory switch.[6] Stochastic models are another approach used in molecular biology modeling. Processes like genetic expression and regulation, signal transduction and cellular reproduction exhibit a stochastic behavior.[7] A stochastic kinetic analysis of the phage λ lysis-lysogeny decision circuit resulted in statistics on regulatory outcomes.[8] Many literature reviews presenting these modeling approaches and others have been published, and some books introducing the subject are available.[9–12]

Discrete event simulation is another interesting approach for molecular biology modeling. Categorized in this last family of approaches, Petri nets can serve to model, analyze and simulate biological processes. The use of Petri nets in biology was suggested for the first time by Reddy *et al.*, who qualitatively analyzed metabolic pathways.[13] Since then, several types of biological processes have been modeled and simulated with Petri nets, mainly molecular biology systems, but also in epidemic and ecologic modeling.[14–16] Peleg *et al.* assessed Petri nets and ten other types of model from the fields of software engineering, business and biology to evaluate their appropriateness for representing and simulating biological processes.[17] Their conclusions were that the combination of two of the assessed types of model, workflow and Petri net models, was the most suitable notation. Aptness of Petri nets for biological research is also demonstrated in recent articles.[18–23] Furthermore, software tools for molecular biology modeling and simulation based on a Petri net architecture are being developed.[24–26]

Despite the various works with different Petri net approaches, Chen *et al.* emphasized that "they lack unity in their concepts, notations and terminologies. This makes it very difficult for new scientists to understand the potential applications of Petri nets due to the various interpretations presented by different authors."[27] The aim of this paper is to analyze the modeling of biological system with the various types of Petri nets. For the complete methodology underlying each approach, the referenced articles should be consulted in their entirety. In the next section, elements from Petri net theory and the earliest attempts at using them for modeling are introduced. Then, the modeling and simulation of biological process with stochastic, colored and hybrid Petri nets are presented and the characteristics of each approach are discussed. The glycolysis pathway, modeled with three different types of Petri net, is illustrated. The formal definition of each Petri net type and the definitions of some of their properties are included with each presentation. Petri nets have also been used to analyze metabolic pathway

databases.[28] In this kind of application, Petri nets are useful for comparing data from various sources, but this is not within the scope of this paper.

2. Elements of Petri Net Theory and the Earliest Attempts at Biological Modeling

Petri nets were introduced at the beginning of the sixties by Prof. Carl A. Petri as a mathematical modeling tool to express system properties like concurrency, indeterminism, communication and synchronization. The basic Petri net is also called a place/transition net. It is founded on a mathematical formalism of the oriented graph. Petri nets and their changes of state can respectively be transposed into matrices and matricial operations. Petri nets are a network where tokens located in places will initiate transitions according to given conditions that will result in the generation of new tokens in the output places. Petri nets and their various elements have a standardized graphical representation as shown in Fig. 1: places p_1, p_2 and p_3 are circles, transition t_1 is a full rectangle, individual tokens are full dots and arcs are drawn as arrows where a positive integer indicates their weight. The number of tokens in each place is usually associated with variables, which are m_1, m_2 and m_3 in the figure. Places can contain tokens which move from place to place via transitions. Tokens are not distinguishable from one another. An arc always binds a place to a transition. Definition 1 is the formal definition of a Petri net.

Definition 1. The Petri net N is defined by the n-tuple (P, T, Pre, $Post$, M) where:

$P = \{p_1, p_2, \ldots, p_u\}$, a finite set of places where $u > 0$;
$T = \{t_1, t_2, \ldots, t_v\}$, a finite set of transitions where $v > 0$;
$P \cap T = \oslash$, meaning that the sets P and T are disjointed;

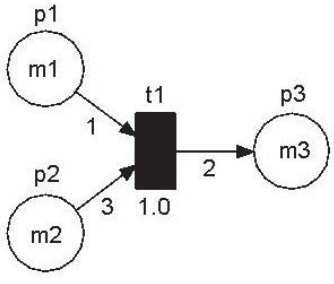

Fig. 1. Places p_1 and p_2 are input places and place p_3 is an output place of the transition t_1. The token contents of places p_1, p_2 and p_3 are m_1, m_2 and m_3 respectively. The weight constants 1 and 3 on the arcs going out of places p_1 et p_2 and the value 1.0 attached to transition t_1 mean that t_1 can fire if $m_1 \geq 1$ and $m_2 \geq 3$, and that the firing delay is 1.0 time unit (in the case of a timed net). When t_1 is fired, one token is removed from p_1, three tokens are removed from p_2 and two tokens are added to p_3.

$Pre = P \times T \rightarrow \mathbb{N}$, is the input incidence mapping (weights of the arcs going from places to transitions) and where $\mathbb{N}$ is the set of natural numbers;

$Post = P \times T \rightarrow \mathbb{N}$, is the output incidence mapping (weights of the arcs going from transitions to places);

$M = P \rightarrow \mathbb{N}$, is the marking of the net which is a vector of u components $(m_1, m_2, \ldots, m_u)$, where m_i is the number of tokens contained in the place p_i. M_0 is the initial marking.

The marking M of the network gives the state of the Petri net. It is a vector indicating the number of tokens $M(p)$ at each place p. When a transition is fired, there is a change in the state of the net, and consequently a modification of the marking. A firing can occur when all the input places of a transition contain the minimal number of tokens defined by the *Pre* relation. In other words, when the number of tokens of all input places is greater than or equal to the weight of the arc linking them to a transition, this transition can fire. Then, the tokens are consumed by the transition and withdrawn from the input places, just as other tokens are created and added to the output places of the same transition. The number of tokens created is specified by the *Post* relation. Figure 1 illustrates how a Petri net transition works. For a more formal and complete coverage of traditional Petri nets and an analysis of their structural properties, consult Reisig's introduction on the subject.[29] Many extensions have been added to the initial model, the purpose of which is to transform models into a more compact form, to elevate the abstraction level or to give Petri nets new capabilities. Some of them have been used for modeling and simulation in biology, and these will be briefly presented in the following sections. The similarities between modeling in molecular biology and Petri net theory are thoroughly discussed.[23]

Traditional Petri nets were originally suggested for biological pathway modeling by Reddy *et al.*, and the bridging of molecular species and chemical reactions with Petri net places and transitions was achieved for the first time by them.[13] The association of places with molecular species and transitions with chemical reactions is used for all types of Petri net model presented in this review. However, special situations necessitate more than one place for one species, for example, when distinguishing between an enzyme in an activated or a deactivated state, or a metabolite in various sites of the cell. The number of tokens indicates the quantity of substance and it corresponds to a predefined measure unit according to the scale of the model, such as the exact number of molecules, mole, millimole, etc. Reddy demonstrated that the Petri net approach was an appropriate tool for a preliminary qualitative analysis of biopathways. Behavioral and structural properties of Petri nets, like liveness, boundedness and invariants were used to identify some characteristics of models (see Definitions 2 to 6). This analysis approach was applied to the erythrocyte pentose phosphate pathway and to the main glycolytic pathway (see Fig. 2 for model and Tables 1 and 2 for symbols definition).[30] The analysis of these pathways showed boundaries for certain molecular species, conservation

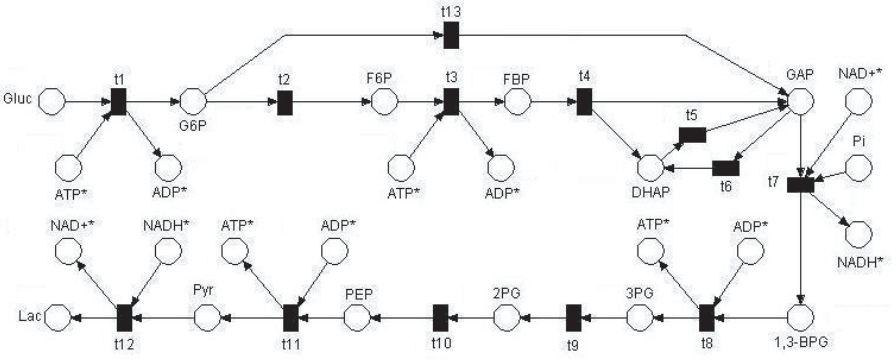

Fig. 2. The Petri net model of the combined metabolism of the glycolytic and pentose phosphate pathways of an erythrocyte cell transforming glucose into lactate (pentose phosphate pathway is abstracted by transition t_{13}). Places labeled with an asterisk are, in fact, one place only, which has been divided for reasons of clarity. ATP, for example, has only one place. See Tables 1 and 2 for elements definition. Model inspired after Fig. 8 from Reddy *et al.*[30]

Table 1. Mapping between metabolites in the pathway and places in the Petri net models of Figs. 2, 3 and 6.

	Metabolite/Compound	
Abbreviation	Marking Variable Associated to Concentration	Name
Gluc	m_1	Glucose
G6P	m_2	Glucose-6-phosphate
F6P	m_3	Fructose-6-phosphate
FBP	m_4	Fructose biphosphate
DHAP	m_5	Dihydroxyacetone phosphate
GAP	m_6	Glyceraldehyde-3-phosphate
1,3-BPG	m_7	1,3-Biphosphoglycerate
3PG	m_8	3-Phosphoglycerate
2PG	m_9	2-Phosphoglycerate
PEP	m_{10}	Phosphoenolpyruvate
Pyr	m_{11}	Pyruvate
Lac	m_{12}	Lactate
ATP	m_{13}	Adenosine triphosphate
ADP	m_{14}	Adenosine diphosphate
NAD+	m_{15}	Nicotinamide adenine dinucleotide, oxidized form
NADH	m_{16}	Nicotinamide adenine dinucleotide, reduced form
Pi	m_{17}	Orthophosphate, ionic form

properties, regenerative reactions and situations leading to a deadlocking of the system.

Definition 2. *Reachability.* A marking M is reachable if it can be reached from the current marking M_i in a finite firing sequence.

Table 2. Mapping between reactions in the pathway and transitions in the Petri net models of Figs. 2, 3 and 6.

Index	Enzyme/Reaction
t_1	Hexokinase
t_2	Phosphoglucose isomerase (forward reaction)
t_3	Phosphofructokinase
t_4	Aldolase
t_5	Triosephosphate isomerase (forward reaction)
t_6	Triosephosphate isomerase (reverse reaction)
t_7	Glyceraldehyde-3-phosphate dehydrogenase
t_8	Phosphoglycerate kinase
t_9	Phosphoglycerate mutase
t_{10}	Enolase
t_{11}	Pyruvate kinase
t_{12}	Lactate dehydrogenase
t_{13}	Pentose phosphate pathway abstraction
t_{14}	Phosphoglucose isomerase (reverse reaction)

Note: The reactions associated with transitions t_{12} to t_{43} of the model of Fig. 6 are identified in its caption.

Definition 3. *Boundedness.* A place is bounded with bound k, if the token count does not exceed k for any reachable marking M of the net. A Petri net is k-bounded if each place is k-bounded.

Definition 4. *Liveness.* A transition is potentially firable if there exists a sequence of transition firings leading to a marking in which the transition is enabled. A transition is live if it is potentially firable for all reachable marking. A transition is dead if it is not potentially firable at the marking M; so if the Petri net enters marking M, the dead transition cannot fire any more.

Definition 5. *S-invariant.* If C is the incidence matrix corresponding to the result of *Post* — *Pre*, then S-invariants are the solutions to the equation $Cy = 0$. The non-zero entries in the vector y constitute the set of places whose total token count does not change with any firing sequence. It is a conservation rule.

Definition 6. *T-invariant.* If C is the incidence matrix corresponding to the result of *Post* — *Pre*, then T-invariants are the solutions to the equation $C^T x = 0$, $x \geq 0$. The solution vector x is the set of transitions that have to fire, from some marking M, to return the Petri net to the same marking M. It is a regenerative rule.

Hofestädt modified Petri net formalism in order to adapt it to metabolic network modeling.[31] The objective was to model metabolic processes in a natural way by developing a new formal graphical representation based on Petri nets. Places and transitions were specialized: places were either metabolite or enzyme, and transitions were either biosynthesis, cellular communication or protein biosynthesis. Unfortunately, the advantage given by this increase in modeling power is of no use for solving the simulation issues that arose at that moment. Later work of

Hofestädt and Thelen suggested a solution to this problem: the application of a Petri net extension corresponding more to the biological context.[32] This extension is the self-modified Petri net, first defined by Valk and now known as the functional Petri net (see Definition 7).[33] The main feature of this augmented formalism is the possibility of assigning to Petri net arcs an equation using marking variables instead of a positive integer. The result is that network marking dynamically modifies the weight of the arcs. In the case of a biochemical process model, this feature is particularly useful for simulating the influence of the variation in concentration on the kinetic rate of biocatalytic reactions. In other words, concentrations, represented by the number of tokens in the net, are variables for the functions that define the weight of the arcs. Thus, the reaction rate of a transition is modified according to the concentration of the various substances involved. Hofestädt and Thelen suggested that quantitative simulations with self-modified Petri nets would help detect metabolic bottlenecks in some defective processes.

Definition 7. The functional Petri net N is defined by the n-tuple (P, T, Pre, $Post$, V, M) where:

(P, T, Pre, $Post$, M) is a Petri net as described in Definition 1;

$V = \{g_a(m_1, \ldots, m_u), a \in Pre \cup Post \mid g : p_1 \times \cdots \times p_u \to \mathbb{N}\}$, a set of functions assigned to arcs of the net using its marking $(m_1, \ldots, m_u)$ as parameters.

3. Stochastic Petri Nets

The random nature of molecular interactions at low concentration has been observed in several experiments. However, the Kolmogorov equations of the stochastic model corresponding to a biological system rapidly become impossible to solve analytically. Goss and Peccoud used stochastic Petri nets (SPN)[34] as a tool for biological modeling of stochastic models.[7] They implied that the Petri net formalism and its modeling power can reduce model implementation delays. With their model, they successfully analyzed the stabilizing effect of the ROM protein on the genetic network controlling the replication of ColEl plasmid replication.[35] More recently, the response of transcription factor σ^{32} to a heat shock and the intracellular kinetics of a viral invasion have been studied through simulation of SPN models.[36,37]

In the SPN model of a system composed of molecular interactions, each place corresponds to a particular molecular species. Tokens represent molecules and transitions between places are chemical reactions involving reactants (input places) and products (output places). At any time, the marking of the system indicates the number of molecules of each species involved. The values of arcs originating from input places and ending at output places are the equivalent of stochiometric coefficients. As in traditional place/transition nets, if the number of tokens at input places is higher than the weight of all the input arcs of a transition, this transition can fire. In molecular terms, the firing of a transition means that a chemical

reaction is occurring. The particularity of SPN is that the firing of a transition is not instantaneous. There is a delay following a probabilistic distribution, thus the delay is a random variable (see Definition 8). In SPN biological models, this delay is interpreted as the reaction rate, and it is given by the weight function of the corresponding transition. The delay mean time is obtained by the transition reaction rate, which is a function of a stochastic rate constant and the quantity of each molecular species involved as a reactant or a catalyst. This constant takes into account volume, temperature, pH and other environmental factors. It is also related to the deterministic rate of the reaction. Several types of SPN exist, but, in the type discussed here, the same type that is used by Goss and Peccoud for modeling a biological process, the weight function will take into account the marking of the net in order to correctly calculate reaction rates. When the number of molecules is sufficiently large, the stochastic constant of the reaction rate is equal to the deterministic rate.

Definition 8. The stochastic Petri net N is defined by the n-tuple $(P,\ T,\ Pre,\ Post,\ F,\ \lambda,\ M)$ where:

$(P,\ T,\ Pre, Post,\ M)$ is a Petri net as described in Definition 1;

$F = \{F_t, t \in T \mid F_t\colon [0,\infty) \to [0,1]\}$, a set of probability density functions for the net firing delays. Their average is 1 and they are independent of the marking;

$\lambda = \{\lambda_t, t \in T \mid \lambda_t\colon \mathbb{N} \to \mathbb{R}^+\}$, a set of firing rates, which are function of the marking (a set of natural integers) and where each element is associated with a transition t. This rate, a positive real number from the set $\mathbb{R}^+$, is used to calculate the probability density function for the transition t.

Stochastic models are applicable when molecules are considered as a discrete amount. Then, a deterministic change in concentrations, quantified by reaction kinetic rates of the incessant flux of transformation of reactants in products, becomes a random event where reactions are ruled by probabilistic laws. SPN can help build these models from their reaction equations and simulate them. It consequently becomes possible to study a system with the simulation results.

The software Möbius (or UltraSAN in its earlier version) has been used for all the systems modeled with SPN mentioned in this section.[38] This SPN simulation tool — not exclusive to biology — also has a model numerical resolution option. With this tool, the molecular species distribution can be studied and the occurrence probability of certain events can be calculated. For example, in the analysis of the stabilizing effect of the ROM protein on a genetic network, the probability that a cell will divide without having replicated its plasmid was estimated.

4. Colored Petri Nets

The differentiation between categories of tokens when modeling large systems with Petri nets was considered in order to reduce the size of models. Thus, Petri nets were enhanced with this new feature by adding colors. The resulting high-level

net, colored Petri net (CPN), is composed of tokens identified by a color (see Definition 9). With this augmentation of the formalism, it is possible to represent, in the same model, different dynamic behaviors modeled by different token colors.[39] Two research teams modeled and simulated biological processes with CPN. In all projects, Design/CPN software, consisting of a set of edition and simulation tools for CPN, has been used.[40] However, each team had a different modeling goal and the conceptual meaning of the token colors was not the same.

Definition 9. The colored Petri net N is defined by the n-tuple (P, T, Pre, $Post$, C, M) where:

(P, T, Pre, $Post$, M) is a Petri net as described in Definition 1 and the tokens of M are identified by a color;

$C = \{C_1, C_2, \ldots\}$, a set of colors. The incidence mappings Pre and $Post$ are functions of the token colors.

Genrich *et al.* modeled an enzymatic reaction with a colored Petri net transition.[41] This transition is connected to places representing substrates like re-actant, product, enzyme and inhibitor. In this model, tokens are identified by two colors, one associated with the substance name and the other with its concentration. The CPN used for this modeling also has functional features because an execution model is called upon, after every firing of the transition, to calculate and modify substrate concentration. These reaction rate calculations are performed according to the Michaelis–Menten biochemical equation, augmented by an additional term for the free reaction energy. The specific constants associated with each enzyme needed for these calculations are extracted from the BRENDA biochemical database.[42] This transition is, in fact, a sub-model integrated into the glycolysis and citric acid metabolic models. A chain of enzymatic reactions constitutes the metabolic network to be quantitatively simulated. Another interesting part of the Genrich *et al.* paper is to propose rules for automatic pathway identification from databases, after which the pathways are modeled as Petri nets for simulation purposes.

Although the CPN was used in their work to model a metabolic system, Heiner, Koch and Voss proceeded differently to accomplish a qualitative analysis of steady states in pathways.[43,22] The objectives of this approach are compatible with the work of Reddy, but the modeling power and the communication capacity of the model are enhanced by the addition of colors. They refined the initial Reddy model of glycolysis and pentose phosphate pathways in an erythrocyte cell by the inclusion of reversible reactions and flux modes (see Fig. 3 for model). Unlike to Genrich, the intention in using colors was to separate branches of a metabolic pathway and to differentiate molecules of the same species (thus, tokens of the same place) according to their origin and destination reaction. The analysis of the invariants of the CPN model found a preservation law for the amounts of all metabolites in the system and confirmed regenerative reactions and their partial order.

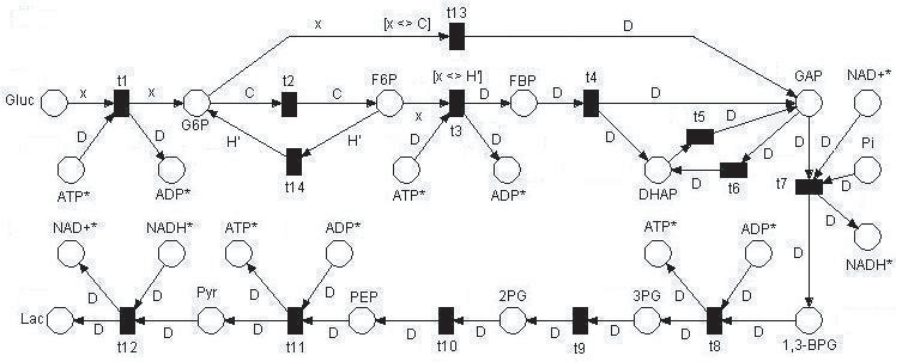

Fig. 3. Colored Petri net model of the glycolysis and pentose phosphate pathways of an erythrocyte cell. The colors used in this model are C, D and H′. The variable X means any color. Three flux modes are represented in the original model. The first mode is glycolysis, associated with the guard [X <> C] of t_{13}, which means that tokens of color C cannot travel through the pentose phosphate pathway. The two other modes are combined in the pentose phosphate pathway (all details are not shown), one of them is associated with the guard [X <> H′]. Model inspired after Fig. 3 from Heiner *et al.*[43]

5. Hybrid Petri Nets and Supplementary Extensions

An intuitive way for representing a molecular species concentration is with tokens of a continuous nature instead of a discrete nature. The hybrid Petri net (HPN) offers this possibility with a new continuous type of places and transitions (see Definition 5).[44] In HPN, discrete places and transitions, with their number of tokens represented by integers and their possible firing delay, are unchanged. But, in the new continuous places, tokens are replaced by a non-negative real number called a mark, and a variable called speed is assigned to the new continuous transitions. The continuous transition speed is a rate of quantity transformation from input places to output places. Thus, the modeling of metabolic reactions and genetic regulation, usually performed with ODEs, can now be accomplished with hybrid Petri nets. Figure 4 explains how a continuous HPN transition operates, and Fig. 5 shows the HPN elements graphically. Matsuno *et al.*[45] and Chen and Hofestädt[18] have demonstrated the feasibility of modeling biological systems with HPN.

Definition 10. The hybrid Petri net N is defined by the n-tuple (P, T, Pre, $Post$, h, M) where:

(P, T, Pre, $Post$, M) is a Petri net as described in Definition 1, where M is a combination of integers for the number of tokens in discrete places and of real numbers for the mark of continuous places;

h: $P \cup T \rightarrow \{D, C\}$, called a hybrid function, indicates for each place and transition, if it is discrete ($h(p_i) = D$ and $h(t_j) = D$) or continuous ($h(p_k) = C$ and $h(t_1) = C$);

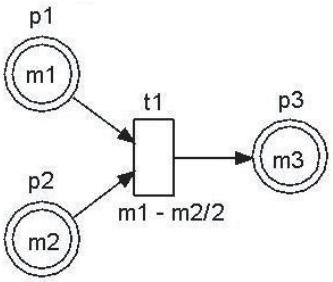

Fig. 4. Places p_1, p_2 and p_3 are continuous places having content m_1, m_2 and m_3 respectively. The function $m_1 - m_2/2$ is assigned to the continuous transition t_1 as its firing speed, t_1 can be fired if $m_1 > 0$ and $m_2 > 0$. The contents of p_1 and p_2 are consumed with the speed $m_1 - m_2/2$, and the content of p_3 increases with the same speed when the transition t_1 is fired.

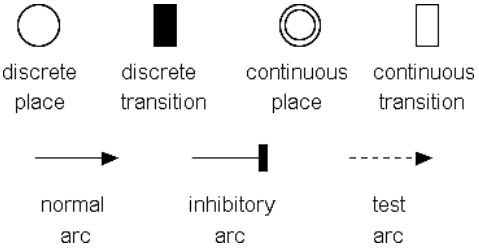

Fig. 5. Graphical representations of the elements of hybrid Petri nets.

A delay d_t is assigned to all discrete transitions and a speed v_t is assigned to all continuous transitions.

Matsuno *et al.* modeled the genetic switching mechanism of the λ phage with HPN[45] by using the Visual Object Net++ tool.[46] Part of their work demonstrated that the HPN approach constitutes a more intuitive modeling tool than the ODE, because of its graphical notation as much as for its power as a simulation tool. The variations in expressed protein concentration, resulting from the simulations, correspond to the biological data on that system.

To be able to adapt the HPN modeling approach to biological processes more accurately, functional Petri net properties have been added to create a new extension: the hybrid functional Petri net (HFPN).[24] This twinning allows a dynamic adaptation during the execution of the Petri net like that in the Genrich *et al.* work.[41] The attribution of a value to the net arcs makes it possible to model biochemical reaction rate equations like the Michaelis–Menten equation. Moreover, two new arc types, different from "normal" arcs, are included in the HFPN to model biological aspects. Firstly, inhibitory arcs model the inhibition function of molecules in some reactions. An inhibitory arc with weight r enables the transition to fire only if the content of the place at the source of the arc is less than or equal to r.

Secondly, test arcs verify the presence of content at its source place when the related transition fires without consuming anything. Unlike inhibitory arcs, some content is necessary in a place connected with a test arc to a transition about to fire. With the inhibitory arc, the repressing function of an operator on gene transcription can be represented. With the test arc, the action of an enzyme in a metabolic reaction, where the enzyme is required but not consumed by the reaction, can be modeled. From the HFPN architecture, the Matsuno team has developed biosimulation software for biologists called Genomic Object Net (GON).[25,26,47]

Many biological processes and systems have already been modeled and simulated using HFPN. Some processes are related to genetic regulation: the λ phage genetic control mechanism,[45] circadian rhythms in *Drosophila*[19] and the control mechanism of the lac operon of *E. coli*.[20] Others are metabolic networks: the glycolytic pathway[32] (see Fig. 6) and the urea cycle.[18] The transduction signal system of apoptosis induced by the Fas ligand was also modeled with HFPN.[19] The pattern formation by a multicellular system due to interactions between cells with Delta-Notch signaling was simulated.[48] In this last experiment, a cellular boundary formation in *Drosophila* and other abnormal patterns could be analyzed with simulation results, corroborating observations from laboratory experiments.

The scientists who originated the HFPN architecture are continuing its elaboration by planning the incorporation of extensions to enhance its modeling power and simulation precision.[19] To achieve this, more complex information like the

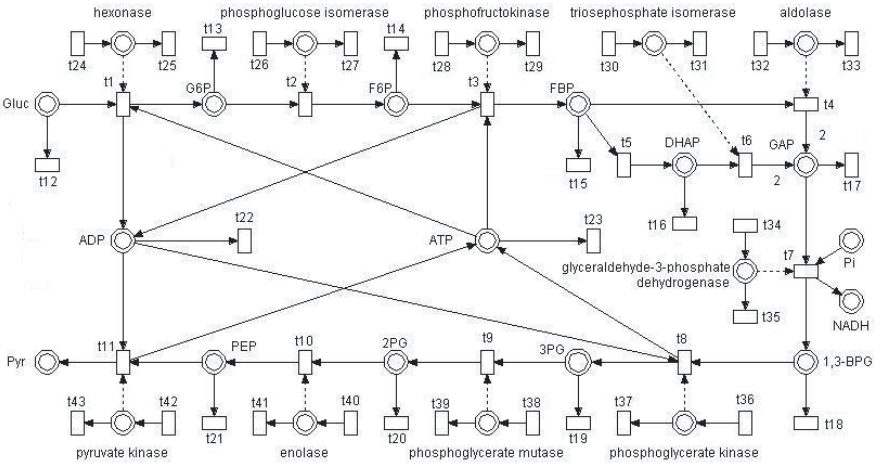

Fig. 6. Hybrid functional Petri net model of the glycolytic pathway and lac operon gene regulatory mechanism. Transitions t_{12}–t_{23} are the natural degradation of substrates and their firing speed is given by the formula $mX/10000$ where $X = 1, 2, \ldots, 10, 12, 13$. Production rate of enzymes $(t_{24}, t_{26}, \ldots, t_{42})$ has a speed set to 1 and their degradation rate $(t_{25}, t_{27}, \ldots, t_{43})$ has a speed given by the formula (*enzyme concentration*)/10. The reactions in the main pathway $(t_1, t_2, \ldots, t_{11})$ have the Michaelis–Menton equation for speed: $\frac{V_{\max} mX}{K_m + mX}$ where $X = 1, 2, \ldots, 10, V_{\max}$ is the maximum reaction speed and K_m is a Michaelis constant. Model inspired after Fig. 9 of Ref. 20.

localization of biological objects or intercellular interactions at the molecular level must be incorporated into the models. One of these extensions makes it possible for a Petri net place to have various data types, like integers or real numbers, vectors and strings. Also, a conversion module implementing the procedure will be capable of converting biological models represented with ODEs in the E-CELL software format to HFPN models.[49]

6. Discussion

To make the modeling power of Petri nets richer and to adapt them to diverse problems, several extensions have been added to Professor Petri's theory. However, enhancements given by the extensions presented in this review do not offer the same modeling and simulation possibilities. Colored Petri nets (CPN) were created to diminish model size and so allow the models to manage more information without rendering its structure too complex. Stochastic Petri nets (SPN) are a specialized member of the timed Petri net family where firing delays are random variables. The need to represent discrete and continuous quantities in the same model motivated the development of hybrid Petri nets. Finally, the dynamic modification of arc weight through net marking is possible with functional Petri nets. Other extensions were developed to solve other problems. With the variety of Petri net types available, the context of use of the various extensions for biological process modeling and simulation can be called into question.

To analyze each approach and its biological modeling possibilities, it is pertinent to recall the compromise discussed by Reddy *et al.* between the modeling and simulation power and the decision-making power of Petri net.[30] According to this compromise, the addition of modeling power to a Petri net with extensions will decrease its analytic capabilities. Indeed, augmenting modeling power amplifies the complexity of the determination of some properties, even to the point of indecision. For example, inhibitor arcs expand the richness of concepts expressed by a model, but greatly complicate its mathematical analysis. Thus, it is important to choose the Petri net extension that will be used for modeling judiciously, in accordance with the objectives to be attained.

In the literature, there are two categories of goals of Petri net biological modeling: **qualitative** and **quantitative** analysis. One can either learn more about the properties of the system under study with a qualitative approach or study the system dynamics with simulation. When one wants to analyze a complex system of biochemical reactions, for example by identifying invariants, the presence of boundaries or the liveness in the system model, the Petri net extension chosen must enable those properties to be determined. By contrast, one may want to study the model's behavior by simulating it and thus obtain concentration graphs, and/or observe the achievability of a steady state, without any concern for property decidability. In that case, modeling power is more important. Thus, it is possible to identify modeling goals for each Petri net extension (see Table 3).

Table 3. Biological modeling goals of Petri net extensions.

Petri Net Extension	Modeling Goal	Analysis Type	Process Type	Available Software	References
Colored	Analysis of biological system properties	Qualitative	Metabolic	Design/CPN	22, 42, 44
Stochastic	Simulation of biological systems with low concentrations	Quantitative	Any biological stochastic process	Möbius	7, 36, 37, 38
Hybrid functional	Simulation of biological systems	Quantitative	Metabolic, Regulatory networks, Signal transduction	Genomic Object Net	18, 19, 24, 25, 26, 46, 49

As was the case for the earliest modeling attempts with Petri net by Reddy *et al.*, where a qualitative analysis was performed using place/transition nets, CPN can give some insights into a biological system.[43] Thus, a rigorous preliminary analysis can guide the elaboration of experiments when quantitative data is missing. Token color does not reduce the decidability power of Petri nets because it is always possible to convert each CPN mode into a traditional net. One advantage of using CPN is the possibility of discriminating metabolites on the basis of their chain of reactions in a model. However, the size of the model must not exceed a certain limit because if the model complexity is too great, the state space will explode and its complete exploration and analysis will be impossible. Consequently, the analysis of complex systems is still a difficult task. Until now, this method has only been used on classic metabolic systems (like the Krebs cycle or glycolysis) to demonstrate its potential, but has also been incorporated in algorithms to find new metabolic pathways between two compounds.[28] Little information is needed to perform a qualitative analysis: the stoichiometry and the reversibility of the system reactions.

Hybrid and stochastic net attributes are intended for simulation, and token activity in the model is the main aspect considered in reproducing the behavior of a system. Wanting to model a system in order to quantitatively simulate it is in accordance with the modeling goals of Hofestädt and Thelen.[32] The criterion of choice between the stochastic and the hybrid extensions is the nature of the system to be modeled. If, for example, a model deals with a small number of molecules, such that their individuality has to be taken into account, its stochastic nature has to be represented and SPN are appropriate. By contrast, for models where the number of molecules is high enough to be represented in a satisfying way as a continuous quantity or as a concentration, HPN is the appropriate modeling approach. It is interesting to note, however, that, if the discrete transition delay in HPN is

modified to become a random variable, then it is possible to fuse HPN with SPN. Matsuno *et al.* mentioned that their architecture could be easily modified to allow this blending.[45] It is also important to recall that the twinning of the functional and hybrid extensions, as in the HFPN, facilitates the modeling of metabolic reactions.

To accomplish a quantitative analysis of a biological system, kinetic parameters like reaction rates have to be available. This kind of analysis provides means to test new hypothesis and to evaluate the impact of the variation of conditions in which a system evolves. It has proven its usefulness in several biology research projects. As mentioned above, the stabilizing effect of the ROM protein on the genetic network controlling the replication of ColE1 plasmid replication was successfully analyzed with a SPN model[35] and a cellular boundary formation in *Drosophila* could be analyzed with simulation results of a HFPN model, corroborating observations from laboratory experiments.[48]

The work of Genrich *et al.*[41] is an exception to the classification shown in Table 3. In their model, the color of the tokens is associated with the concentration of the molecular species in order to make a quantitative analysis. This demonstrates that the classification is not absolute. However, in addition to the colored extension, the Petri net type used in that model also includes an executable component which is called upon at every transition firing to calculate the concentration variations according to the Michaelis–Menten equation. Despite the validity of this approach, it is less intuitive and harder to implement than an approach using HFPN.

Using the Petri net for biological modeling offers many advantages. First, two Petri net properties, identified by Reddy *et al.*, are pertinent to the modeling of biochemical systems: extendibility and abstraction.[30] These two features are related to model hierarchization. Extendibility is the property of adding new sections to a net — for example, when supplementary information becomes available to complete a model, or when one wants to combine two complementary models — without having to considerably modify the structure of the resulting model. Abstraction is the property of neglecting the modeling of some aspects which do not concern the system under study by representing the sub-model by a transition. An example of abstraction is the transition t_{13} of the Figs. 2 and 3 representing the pentose phosphate pathway.

Second, many theoretical elements of Petri nets with a mathematical basis are useful as a preliminary analysis tool for biological pathways. Zevedei–Oancea and Schuster thoroughly discuss this.[23] Oliveira *et al.* developed and defined rigorously a computational approach based on Petri nets to identify interesting sub-circuit pathways in biochemical networks and applied their methodology to the Krebs cycle.[50,21] Petri net invariants can be associated with flux modes and conservation relations. Furthermore, special sets of places known as siphons and traps can be identified. The places constituting a siphon stay empty once they have no tokens. At the other extreme, places forming a trap cannot lose one token when they reach a certain marking. Traps and siphons are of interest in biochemical modeling

because these notions can be associated with the storage and consumption of system resources. Finally, the liveness of a net or its deadlocked condition are also properties which provide information about the biological system.

Third, biologists can easily model a biological system with the Petri net, and study it with the simulation capabilities of Petri net tools. The graphical aspects of the Petri net are quite similar to biochemical network representation, and this gives superior communication ability to models and facilitates their design. Moreover, the Petri net is readily comprehensible and necessitates little related knowledge on the part of biologists. Its mathematical basis makes it possible to accomplish complex simulations and to visualize results. The development of software based on the Petri net and specific to biology, like the Genomic Object Net tool,[25,26,47] and the proposal of a data exchange format for models, the "Biology Petri Net Markup Language" (BioPNML)[27,51] leave few obstacles for the adoption of a Petri net approach by biologists. A powerful analysis and simulation environment can be implemented from this modeling technique to study hundreds, even thousands, of interconnections formed by the various genetic and metabolic networks in the cell.

Several types of formal representation other than Petri nets are required to model biological processes. It was demonstrated that ordinary differential equations can be substituted by HFPN, but other biological phenomena involving spatial modeling, like diffusion, or molecular motions modeling, like molecular motors, do not have an equivalent in Petri net modeling. Because of the potential extensibility of Petri net formalism, it is possible to think that scientists will go beyond these modeling limits. Thus, we could also say that these biological phenomena do not have an equivalent in Petri net modeling *yet.*

7. Conclusion

In this paper, analysis, modeling and simulation of molecular biology systems using Petri nets have been presented and an overview of various approaches using colored, stochastic, hybrid and functional Petri nets was made. The modeling goal of each approach was identified, thus providing a starting point to interested new users.

Petri net is a formalism with many advantages for biologists. It has analytical and simulation capabilities which provide means to test hypotheses and gather information that might help the elaboration of experiments. As we learn more about metabolic pathways, gene regulatory networks and signalling pathways, powerful modeling tools like Petri net will be needed to understand the complexity of living systems.

Acknowledgments

We would like to thank Prof. Hanifa Boucheneb for her insightful comments and the two anonymous reviewers for their useful remarks about this article. This work was supported in part by NSERC grant A-0141 and a Canada Graduate Scholarship.

References

1. Kitano H, Computational systems biology, *Nature* **420**:206–210, 2002.
2. Heinrich R, Schuster S, The modelling of metabolic systems. Structure, control and optimality, *Biosystems* **47**:61–77, 1998.
3. Mendes P, GEPASI: a software package for modelling the dynamics, steady states and control of biochemical and other systems, *Comput Appl Biosci* **9**(5):563–571, 1993.
4. Tomita M, Hashimoto K, Tabahaski K, Shimizu T, Matsuzaki Y, Miyoshi F, Saito K, Tanida S, Yugi K, Venter JC, Hutchison C, E-CELL: Software for whole cell simulation, *Bioinformatics* **15**:72–84, 1999.
5. Kauffman SA, *The Origins of Order: Self-Organization and Selection in Evolution*, Oxford University Press, New York, 1993.
6. Yuh CH, Bolouri H, Davidson EH, *Cis*-regulatory logic in the *endo 16* gene: switching from a specification to a differentiation mode of control, *Development* **128**:617–629, 2001.
7. Goss PJE, Peccoud J, Quantitative modeling of stochastic systems in molecular biology by using stochastic Petri nets, *Proc Natl Acad Sci USA* **95**:6750–6755, 1998.
8. Arkin A, Ross J, McAdams HH, Stochastic kinetic analysis of developmental pathway bifurcation in phage lambda-infected *Escherichia coli* cells, *Genetics* **149**: 1633–1648, 1998.
9. de Jong H, Modeling and simulation of genetic regulatory systems: a literature review, *J Comput Biol* **9**(1):67–103, 2002.
10. Endy D, Brent R, Modelling cellular behavior, *Nature* **409**:391–395, 2001.
11. Bower JM, Bolouri H, *Computational Modeling of Genetic and Biochemical Net-works*, MIT Press, Cambridge, MA, 2001.
12. Collado-Vides J, Hofestädt R, *Gene Regulation and Metabolism — Postgenomic Computational Approaches*. MIT Press, Cambridge, MA, 2002.
13. Reddy VN, Mavrovouniotis ML, Liebman MN, Petri net representation in metabolic pathways, *Proc First ISMB*, 328–336, 1993.
14. Bahi-Jaber N, Pontier D, Modeling transmission of directly transmitted infectious diseases using colored stochastic Petri nets, *Math Biosci* **185**:1–13, 2003.
15. Gronewold A, Sonnenschein M, Event-based modelling of ecological systems with asynchronous cellular automata, *Ecol Modelling* **108**:37–52, 1998.
16. Sharov A, Self-reproducing systems: structure, niche relations and evolution, *Biosystems* **25**(4):237–249, 1991.
17. Peleg M, Yeh I, Altman RB, Modelling biological processes using workflow and Petri net models, *Bioinformatics* **18**(6):825–837, 2002.
18. Chen M, Hofestädt R, Quantitative Petri net model of gene regulated metabolic networks in the cell, *In Silico Biol* **3**, http://www.bioinfo.de/isb/2003/03/0030/, 2003.
19. Matsuno H, Tanaka Y, Aoshima H, Doi A, Matsui M, Miyano S, Biopath-ways representation and simulation on hybrid functional Petri net, *In Silico Biol* **3**, http://www.bioinfo.de/isb/2003/03/0032/, 2003.
20. Matsuno H, Fujita S, Doi A, Nagasaki M, Miyano S, Towards biopathways modeling and simulation, *Lect Notes Comput Sci* **2679**:3–22, 2003.
21. Oliveira JS, Bailey CG, Jones-Oliveira JB, Dixon DA, Gull DW, Chandler ML, A Computational model for the identification of biochemical pathways in the Krebs cycle, *J Comput Biol* **10**(1):57–82, 2003.
22. Voss K, Heiner M, Koch I, Steady-state analysis of metabolic pathways using Petri nets, *In Silico Biol* **3**, http://www.bioinfo.de/isb/2003/03/0031/, 2003.
23. Zevedei-Oancea I, Schuster S, Topological analysis of metabolic networks based on Petri net theory, *In Silico Biol* **3**, http://www.bioinfo.de/isb/2003/03/0029/, 2003.

24. Matsuno H, Doi A, Hirata Y, Miyano S, XML documentation of biopathways and their simulations in Genomic Object Net, *Genome Inform* **12**:54–62, 2001.
25. Nagasaki M, Doi A, Matsuno H, Miyano S, Genomic Object Net: I. A platform for modeling and simulating biopathways, *App Bioinform* **2**(3):181–184, 2003.
26. Doi A, Nagasaki M, Fujita S, Matsuno H, Miyano S, Genomic Object Net: II. Modeling biopathways by hybrid functional Petri net with extension, *App Bioinform* **2**(3):185–188, 2003.
27. Chen M, Freier A, Köhler J, Rüegg A, The biology Petri net markup language, *Proc Promise 2002*, 150–161, 2002.
28. Küffner R, Zimmer R, Lengauer T, Pathway analysis in metabolic databases via differential metabolic display (DMD), *Bioinformatics* **16**(9):825–836, 2000.
29. Reisig W, *Petri Nets: An Introduction*, Monographs on Theoretical Computer Science, Springer–Verlag, 1985.
30. Reddy VN, Liebman MN, Mavrovouniotis ML, Qualitative analysis of bio chemical reaction systems, *Comput Biol Med* **26**(1):9–24, 1996.
31. Hofestädt R, A Petri net application of metabolic processes, *J System Analysis, Modeling and Simulation* **16**:113–122, 1994.
32. Hofestädt R, Thelen S, Quantitative modeling of metabolic processes, *In Silico Biol* **1**, http://www.bioinfo.de/isb/1998/01/0006/, 1998.
33. Valk R, Self-modifying nets, a natural extension of Petri nets, *Lect Notes Comput Sci* **62**(ICALP), 464–476, 1978.
34. Ajmone Marsan M, Balbo G, Chiola G, Conte G, Donatelli S, Fransces-chinis G, An introduction to generalized stochastic Petri nets, *Microelectron Reliab* **31**(4): 699–725, 1991.
35. Goss PJE, Peccoud J, Analysis of the stabilizing effect of Rom on the genetic network controllin ColE1 plasmid replication, *Pac Symp Biocomput* **4**:65–76, 1999.
36. Srivastava R, Peterson MS, Bentley WE, Stochastic kinetic analysis of the *Escherichia coli* stress circuit using σ^{32}-targeted antisense, *Biotechnol Bioeng* **75**(1):120–129, 2001.
37. Srivastava R, You L, Summers J, Yin J, Stochastic versus deterministic modeling of intracellular viral kinetics, *J Theor Biol* **218**:309–321, 2002.
38. Möbius and UltraSAN, http://www.crhc.uiuc.edu/PER.FORM/.
39. Jensen K, *Coloured Petri Nets: Basic Concepts, Analysis Methods and Practical Use*, Monographs on Theoritical Computer Science, Springer–Verlag (1992).
40. Design/CPN, http://daimi.au.uk/designCPN/.
41. Genrich H, Küffner R, Voss K, Executable Petri net models for the analysis of metabolic pathways, *Int J STTT* **3**(4):394–404, 2001.
42. BRENDA, http://www.brenda.uni-koeln.de:80/.
43. Heiner M, Koch I, Voss K, Analysis and simulation of steady states in metabolic pathways with Petri nets, in *Third Workshop and Tutorial on Practical Use of Colored Petri Nets and CPN Tools*, Univ. of Aarhus, DAIMI PB-554, 15–34, 2001.
44. Alla H, David R, Continuous and hybrid Petri nets, *J Circuits Syst Comp* **8**(1):159–188, 1998.
45. Matsuno H, Doi A, Nagasaki M, Miyano S, Hybrid Petri net representation of gene regulatory network, *Pac Symp Biocomput* **5**:338–349, 2000.
46. Visual Object Net++, http://www.systemtechnik.tu-ilmenau.de/∼drath/visual.htm.
47. Genomic Object Net, http://genomicobject.net.
48. Matsuno H, Murakani R, Yamane R, Yamasaki N, Fujita S, Yoshimori H, Miyano S, Boundary formation by Notch signaling in Drosophila multicellular systems: experimental observations and gene network modeling by Genomic Object Net, *Pac Symp Biocomput* **8**:152–163, 2003.

49. Matsui M, Doi A, Matsuno H, Hirata Y, Miyano S, Biopathways model conversion from E-CELL to Genomic Object Net, *Genome Inform* **12**:290–291, 2001.
50. Oliveira JS, Bailey CG, Jones–Oliveira JB, Dixon DA, An algebraic-combinatorial model for the identification and mapping of biochemical pathways, *Bull Math Biol* **63**:1163–1196, 2001.
51. Chen M, Modelling and simulation of metabolic networks: Petri net approach and perspective, *Proc ESM 2002*, 441–444, 2002.

Simon Hardy is a Ph.D. student at École Polytechnique de Montŕeal, where he also received a B.Ing. and M.Sc. A. in Computer Engineering. He is a member of the Software engineering research laboratory since 2001.

His research interests are biopathways modeling and simulation, software developement for bioinformatic applications and software process.

Pierre N. Robillard Ph.D, P Eng. is chairman of the computer engineering department and professor of Software engineering with École Polytechnique, Montŕeal, Canada, where he leads the Software Engineering Research Laboratory. His research interests include bioinformatics, software process, software cognition and software quality assurance.

He has been involved in various industrial software development tool projects, including Petri net simulation engines and software process tools. He has published over 200 papers, conferences proceedings and co-authored 3 books on software engineering related topics. The latest book is *Software Engineering*, Addison Wesley, 2003.