Modeling gene regulatory networks using Petri Nets

Jure Bordon, Miha Moškon, Miha Mraz

University of Ljubljana,
Faculty of Computer and Information Science,
Tržaška cesta 25, Ljubljana,
Slovenia,
E-mail: jure.bordon@fri.uni-lj.si

Abstract

Petri nets (PNs) are a qualitative approach to modeling gene regulatory networks (GRNs) that are capable of characterising the dynamical properties of complex systems while also describing their structure with PN graphs. In this paper we show how to construct basic building blocks for GRNs using PNs. We present a solution to a few known weaknesses of using conventional PNs for description of GRNs by using Coloured Petri nets (CPNs), an extension of the ordinary PNs. We show that by using CPNs we can create building blocks which describe the dynamical properties of GRNs more accurately. We also consider solving the problem of undefined kinetic data with the use of fuzzy logic methods.

1 Introduction

Synthetic biology is a rapidly evolving field that has shown high potential in recent years. Its state of the art methods allow us to construct almost arbitrary DNA sequence and therefore define the behaviour of a target biological system. Synthetic biology methods have also shown their potential in information processing, i.e. in the construction of biological information processing systems, which have found many application in various fields such as agriculture, medicine and pharmacy [2].

First step in designing a biological information processing system is similar to its electronic counterpart, the construction of logical scheme that defines its functionalities. Logical scheme is afterwards translated into a biological design. When we are dealing with gene expression based logic, biological design is based on the scheme representing genes and influences among them, i.e. gene regulatory network (GRN). Building blocks used are strictly dependant on the target host of information processing systems, which defines the potential interaction rules and chemical species. Because *in vivo* implementation of such systems is time-consuming and costly, we evaluate the preliminary behaviour and dynamics *in silico*, with the aid of various computer models.

This paper is organized as follows. In Section 2 we present the most recent approaches to GRN modeling. In Section 3 we give the prerequisites based on which our further approach is based: the basics of gene expression, brief theory of Petri Nets and known weaknesses of using ordinary PN for GRN modeling. We consider the use of Coloured Petri Nets to solve these weaknesses and present preliminary results in Section 4. We draw conclusions of our current research and present directions for future work in Section 5.

2 Related work

While there are many approaches to GRN modeling [1, 4, 8], quantitative mathematical models based on system of ordinary differential equations (ODEs) seem to be the most common ones. With these models detailed analysis of the systems' dynamics can be performed on one hand. On the other hand accurate kinetic data must be known for their establishment. The problem lies in the fact that most of these parameters are often hard or even impossible to determine [9]. Construction of exact ODE models is therefore often under question. Recently, approaches that describe the system dynamics qualitatively and are capable to characterise the dynamical properties of complex systems without relying on accurate knowledge of kinetic data are being studied in the context of biological systems modeling [3]. Petri Nets (PNs) present one of the most promising modeling tools emerging in the field of qualitative modeling [11]. We can describe the system behaviour by using PN formalisms and its structure with a PN graph. While basic principles of GRNs can be described by using ordinary PNs, different extensions of PNs give us even greater toolset for more complex description of system dynamics [3].

Extensions of PNs are customized for the problem set they are trying to solve. Hybrid PNs have both continuous and discrete elements, which help in describing continuous system processes [10]. For a more quantitative approach Stochastic PNs (SPNs) are used [5]. We

propose using Coloured Petri Nets (CPNs) for biological systems modeling [6, 7].

3 Basics of gene expression and its description with Petri Nets

Organism behaviour is defined by its hereditary material located within its genome. Genome is encoded with two DNA molecules, intertwined into a double helix. Each DNA molecule is a nucleobase sequence of arbitrary length. It connects to its counterpart DNA molecule via hydrogen bonds between complementary nucleobases, called base pairs, forming a double helix. Genome is therefore a sequence of base pairs. It can be divided into smaller subsequences (genes), which represent the basic unit of heredity. Each gene has a coding sequence, a promoter and regulatory sequence. Coding sequences define the protein that certain gene produces, while the promoter and regulatory sequences define the conditions for the expression of the protein.

Gene expression is only possible if *RNA polymerase* is present and the conditions for expression, defined by regulatory sequence, are met. Regulatory sequences can contain binding sites for proteins (*transcription factors*), which are divided into *activators* and *repressors*. Presence of an activator induces the expression of the gene, while presence of a repressor inhibits it. Gene expression is divided into two phases, *transcription* and *translation*. Transcription starts with RNA polymerase binding to the promotor, and is enhanced by presence of activators and diminished by presence of repressors. The product of transcription is mRNA molecule which is afterwards translated to a protein by ribosomes.

Behaviour of a biological system can be described by gene expression dynamics. Gene expression is not necessarily independent and can be regulated by products of other genes. One gene can regulate expression of many others. On the other hand several genes can regulate expression of only one gene. Genes and their regulation relations form a gene regulatory network (GRN). Gene expression dynamic is described by chemical reactions between different chemical species in the system. Basic reactions are binding and dissociation of transcription factors, transcription, translation and degradation. Dynamics of the established models are based on observed reactions. System state is defined by concentrations of chemical species which are changed by the execution of each reaction.

3.1 Petri nets

Petri Nets (PNs) present a tool for studying system dynamics. PNs enable us to model a system by using a mathematical description, which can be shown as a PN graph. We can evaluate systems behaviour by analysing PNs that serve as a model of this system. They are defined by a 4-tuple $C = (P, T, I, O)$, where $P = \{p_1, p_2, \ldots, p_n\}$, $n > 0$ is a finite set of places and $T = \{t_1, t_2, \ldots, t_m\}$, $m > 0$ is a finite set of transitions ($P \cap T = \emptyset$). The input function I is a mapping from transition t_j to a collection of places $I(t_j)$, known as the input places of the transition. It is defined as $I : T \to P^\infty$. The output function O maps a transition t_j to a collection of places $O(t_j)$ known as the output places of the transition. It is defined as $O : T \to P^\infty$. Places can present multiple input or multiple output of a transition. The numbers of occurrences of the place in the input and output collections are $\#(p_i, I(t_j))$ and $\#(p_i, O(t_j))$, respectively.

Tokens in places serve as a precondition for firing the transition. The number of tokens in each place is arbitrary. Marking of a PN gives us the current position and number of the tokens in each place. Marking can be defined as a function $o(t) : P \to N$, which returns the number of tokens in each place ($N \geq 0, N \in \mathbb{N}$) at time t. Marked PN is defined by a 5-tuple $M = (P, T, I, O, o(t))$. The execution of a PN is controlled by the transition firing. A transition t_j is enabled if for all $p_i \in P$, $o_{p_i}(t) \geq \#(p_i, I(t_j))$ is true. By firing enabled transitions, the tokens are removed from the input places and added to the output places of t_j. This means that for every transition firing, marking changes. New marking is calculated as

$$o'_{p_i}(t+1) = o_{p_i}(t) - \#(p_i, I(t_j)) + \#(p_i, O(t_j))$$

We can denote that in matrix form. Input function is presented as a matrix $D^-[i][j] = \#(p_j, I(t_j))$ and output function as $D^+[i][j] = \#(p_j, O(t_j))$. Transition t_j is enabled if $\mathbf{o}(t) \geq e_j \cdot D^-$ holds, where e_j is unit vector with 1 on location j. New marking as a result of firing t_j is calculated as

$$\mathbf{o}(t+1) = \mathbf{o}(t) + e_j \cdot (D^+ - D^-).$$

3.2 Modeling gene regulatory networks using Petri nets

In order to model biological systems with PNs the meaning of basic PN elements have to be defined, i.e. places, transitions and tokens. We presume that transitions represent chemical reactions. Each reactions has its inputs (reactant) and its outputs (products). Places in PNs represent different chemical species and tokens represent the concentration of these species. The reaction can occur when the concentration of its reactants is greater than zero. This means that there has to be at least one token in the place which represents this reactant. Figure 1 shows a simple chemical reaction, i.e. translation. Place p_1 presents the concentration of *mRNA*. In the reaction of translation, presented by transition t_j, *mRNA* is transformed into a *protein*, presented by p_2. The number of input arrows defines the amount of reactants and the number of output arrows the amount of products, which are created as a result of firing transition t_j. The token in p_1 presents the presence of *mRNA*. Transition t_1 is enabled and upon firing, the concentration of *mRNA* will decrease and the concentration of *protein* will increase. The marking of our PN will change, the token will move from p_1 to p_2.

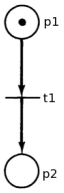

Figure 1: Example of a simple PN that describes translation. Place p_1 presents *mRNA*, place p_2 expressed protein and transition t_1 the reaction of translation.

In formal notation the marking of PN on Figure 1 in time t_0 is $o(t_0) = (1, 0)$. Transition t_1 is enabled, meaning that upon firing, the marking will change to $o(t_0 + \Delta t) = (0, 1)$. Transition firings in ordinary PNs takes infinitesimal little time. In addition, presenting the concentrations with tokens is limited to integer values. We therefore cannot fully describe the change of system state, caused by a chemical reaction. The result of a chemical reaction is not defined only by reactant presence but also by reactant concentration. Figure 2 presents the system state change, as a result of multiple firing of t_j.

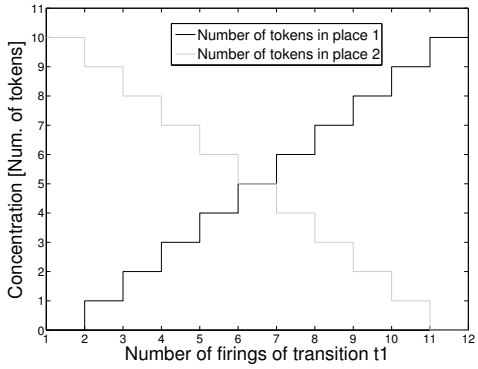

Figure 2: System change as a result of multiple firings of t_j. Initial PN marking is $o(t_0) = (6, 0)$. After six successive firings of transition t_j the marking changes to $o(t + 6\Delta t) = (0, 6)$.

Chemical reactions often reflect the dynamics that can be presented with non-linear-functions. Figure 2 shows that events in ordinary PNs are discrete and changes in concentration increase or decrease evenly with each firing of t_j. System behaviour over time is another important aspect of modeling biological systems. Because time steps in ordinary PNs are infinitesimally small, we are unable to observe the changes as a function of time. We will therefore need to extend the basic PNs, in order to describe the system dynamics in more details.

4 Using Coloured Petri Nets for accurate gene regulatory network description

We propose using CPNs, since they offer solutions to problems presented in Subsection 3.2. Chemical species concentrations in ordinary PNs were described with the number of tokens in certain places. We would like to present concentrations as real numbers. Instead of tokens, concentrations will be presented within the places, as real number values. *Data structures* will be used as carriers of information. Data structures can hold arbitrary data and no longer serve as conditions for transition firing. Data structures and their summary which hold different kinds of data can be interpreted as different colours. For the purposes of biological system modeling we will use two groups of data structures. Data structures in first group will as a result of transition firings contain changes in concentrations of both reactants and products. These changes in concentration depend on concentrations of reactants at the time of firing. Transitions, representing chemical reactions, are therefore more complex and can be presented as functions of input places, which change concentrations in both input and output places. Data structures in the second group will only transfer information about current concentration. Transitions in this case will only serve as buffers and will not present any chemical reactions.

4.1 Redefining the transitions on an example

The definition of transition changes in CPNs, no longer depend on presence of tokens, but rather on state of input places. This enables us to define more complex transitions and set more complex conditions for transition firings. With this extension non-linear dependency can be achieved, since transition can be defined as an arbitrary function. Changes in concentrations are no longer discrete, since data structures can carry data presented with real numbers. By using CPNs we can rebuild the example from Subsection 3.2 using extended transitions and places (see Figure 3). Basic CPN example is built using places P_1 and P_2 which carry the information about current concentrations ($[c_1]$,$[c_2]$), transition t_1, defined by function $f(c_1)$ and data structures "c_1--" and "c_2++", which present the change of concentrations in P_1 and P_2 when t_1 is fired.

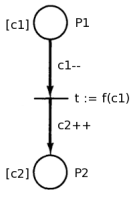

Figure 3: Example of CPN, presenting translation. Places P_1 and P_2 include data about *mRNK* and *protein* concentrations ($[c_1]$,$[c_2]$), respectively. Transition t_1 is a function, representing the chemical reaction of translation. Data structures $c_1 --$ and $c_2 ++$ carry the changes of concentrations as a result of transition firing.

Transition t_1 presents the protein translation, which can be described by the function:

$$f(c_1) = \frac{1}{1 + \frac{c_1}{k_{transl}}}, \quad (1)$$

where c_1 is the concentration in P_1 and k_{transl} kintetic parameter for translation reaction. While we are using extended PNs it is possible to use this function within the t_1 transition. Result of a function in t_1 are changes in concentrations $c_1 - -$ and $c_2 + +$ in places P_1 and P_2. In order to see how system behaves as a function of time, we need to explain how time is interpreted in CPN.

4.2 Time in Coloured Petri Nets

The timings of transition firings can be controlled with the introduction of time into CPNs. CPN dynamics are therefore not described by transition firings, but by the time values that define transition firing timings. Time between two firings is no longer infinitesimally small. This functionality enables us to calculate the results of transitions in time steps, which are now measurable. Actual time evolution of certain species concentrations can be obtained in these manners. Note that transitions in CPNs can include certain functions. E.g., transition t_1 in Figure 3 includes the function $f(c_1)$ which is defined by the Equation 1. Changes in concentration in places for each time step are therefore calculated according to these functions. The results for our scenario are presented in Figure 4.

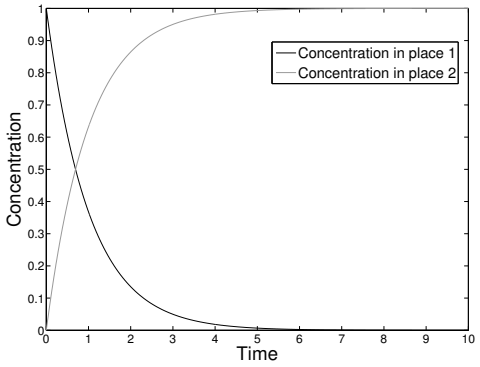

Figure 4: Evolution of system state as a result of function t_1. Initially, the concentration in P_1 is high and the concentration in P_2 zero. Concentration of *mRNA*, presented as c_1 in P_1, decreases with time, while the concentration of *protein*, presented as c_2 in P_2, increases until *mRNA* concentration reaches zero.

5 Conclusion

Our current results of using CPNs for modeling GRNs serve as starting points for future research. We described a basic model of a simple biological system with only one reaction. To describe the dynamics of a complete GRN, we will have to extend existing CPNs in order to support the modeling of more complex biological reactions with the possibility of vaguely defined kinetic rates. Our example transition was presented by a function, that describes the reaction of translation. This function was defined as an equation that used a kinetic parameter. As we have mentioned, these kinetic parameters are often lacking. In future research we intend to tackle the problem of not knowing accurate kinetic data by using fuzzy logic methods.

References

[1] U. Alon, *An introduction to systems biology: Design principles of biological circuits*, Chapman and Hall/CRC mathematical & computational biology series, Chapman & Hall/CRC, 2007.

[2] Ernesto Andrianantoandro, Subhayu Basu, David K. Karig, and Ron Weiss, *Synthetic biology: new engineering rules for an emerging discipline*, Molecular Systems Biology **May** (2006), 1–14.

[3] Claudine Chaouiya, *Petri net modelling of biological networks*, Briefings in Bioinformatics **8** (2007), no. 4, 210–219.

[4] Hana El Samad, Mustafa Khammash, Linda Petzold, and Dan Gillespie, *Stochastic modeling of gene regulatory networks*, International Journal of Robust and Nonlinear Control **15** (2005), 691–711.

[5] Peter J. E. Goss and Jean Peccoud, *Quantitative modeling of stochastic systems in molecular biology by using stochastic petri nets*, PNAS **95** (1998), no. 12, 6750–6755.

[6] Kurt Jensen, *A brief introduction to coloured petri nets*, TACAS, 1997, pp. 203–208.

[7] Kurt Jensen, Lars Michael Kristensen, and Lisa Wells, *Coloured petri nets and cpn tools for modelling and validation of concurrent systems*, STTT **9** (2007), no. 3-4, 213–254.

[8] Mads Kaern, William J. Blake, and J.J. Collins, *The engineering of gene regulatory networks*, Annu. Rev. Biomed. Eng. **5** (2003), 179–206.

[9] Gabriele Lillacci and Mustafa Khammash, *Parameter estimation and model selection in computational biology*, PLoS Comput Biol **6** (2010), no. 3, e1000696.

[10] H Matsuno, A Doi, M Nagasaki, and S Miyano, *Hybrid petri net representation of gene regulatory network.*, Pacific Symposium On Biocomputing **349** (2000), no. 338-349, 341–352.

[11] L. Jason Steggles, Richard Banks, Oliver Shaw, and Anil Wipat, *Qualitatively modelling and analysing genetic regulatory networks: a petri net approach*, Bioinformatics **23** (2007), no. 3, 336–343.