A Multiscale Modeling Framework Based on P Systems

Francisco José Romero-Campero[1], Jamie Twycross[1,2], Hongqing Cao[1], Jonathan Blakes[1], and Natalio Krasnogor[1]

[1] Automated Scheduling, Optimisation and Planning Research Group
School of Computer Science, Jubilee Campus, University of Nottingham
Nottingham NG8 1BB, United Kingdom
[2] Centre for Plant Integrative Biology
Sutton Bonington Campus, University of Nottingham
Nottingham LE12 5RD, United Kingdom
{fxc,jpt,hxc,jvb,nxk}@cs.nott.ac.uk

Abstract. Cellular systems present a highly complex organization at different scales including the molecular, cellular and colony levels. The complexity at each one of these levels is tightly interrelated. Integrative systems biology aims to obtain a deeper understanding of cellular systems by focusing on the systemic and systematic integration of the different levels of organization in cellular systems.

The different approaches in cellular modeling within systems biology have been classified into mathematical and computational frameworks. Specifically, the methodology to develop computational models has been recently called executable biology since it produces executable algorithms whose computations resemble the evolution of cellular systems.

In this work we present P systems as a multiscale modeling framework within executable biology. P system models explicitly specify the molecular, cellular and colony levels in cellular systems in a relevant and understandable manner. Molecular species and their structure are represented by objects or strings, compartmentalization is described using membrane structures and finally cellular colonies and tissues are modeled as a collection of interacting individual P systems.

The interactions between the components of cellular systems are described using rewriting rules. These rules can in turn be grouped together into modules to characterize specific cellular processes. One of our current research lines focuses on the design of cell systems biology models exhibiting a prefixed behavior through the automatic assembly of these cellular modules. Our approach is equally applicable to synthetic as well as systems biology.

1 Introduction

Models in systems biology has been recently classified according to their semantics into *denotational* and *operational* models [6]. Models with *denotational semantics* are the classical approach in modeling cellular systems which uses a

set of equations to describe how the quantities of the different molecular species are related to each other over time. The classical example are ordinary and partial differential equations. In this case the behavior of the system is obtained by approximating numerically these equations. On the other hand, *computational models* have *operational semantics* which describe the behavior of the system using an algorithm or list of instructions that can be executed by an abstract machine. The models developed within this last framework has been termed recently *executable biology* [6]. In this case a more detailed description of the processes producing the behavior of the system is provided.

Several formal computational approaches have been proposed to model cellular systems like Petri nets [10] and process algebra [19]. They mainly focus on system specification at the molecular level: membranes, compartmentalization and cellular colonies are seldom described. This fact makes it difficult to study multicellular systems whose function is determined by molecular interactions.

Membrane computing is a branch of natural computing inspired directly from the structure and functioning of the living cell [14]. It has been applied to cellular modeling as one of the few computational frameworks which presents an integrative approach to multiscale systems ranging from the molecular to the multicellular level. Specifically, it represents the molecular interaction level of living cells using objects or strings and rewriting rules; the compartmental/cellular level using membranes; and the colony level using collections of membranes called membrane structures. The devices of this computational paradigm are referred to as *P systems*. Although most research in P systems focuses on the study of the computational power of the different proposed variants, recently their application as a modeling formalism to cellular systems is emerging [3,4,7,11,17,20,21,15]. In this paper we discuss through a running example the use of P systems as a multiscale modeling framework for cell systems biology models.

The paper is organized as follows. Stochastic P systems for cellular modeling are introduced in Section 2. Section 3 presents the running example used throughout this paper. The modeling principles in P systems are described in Section 4. Modularization in P systems is briefly discussed in Section 5. Finally, conclusions and future work are discussed in Section 6.

2 Stochastic P Systems

The original strategy for the application of the rewriting rules in P systems was based on maximal parallelism and non-determinism [13]. This strategy does not represent the rate at which molecular interactions take place as every object that can evolve according to any rule must evolve in a single computation step, without taking into account that some molecular interactions are more frequent than others. Moreover, the real time evolution of cellular systems is not captured as all the computation steps are assumed to be of the same time length, neglecting the fact that some molecular interactions are faster than others.

Different strategies for the application of the rewriting rules in P systems have been studied [5,8]. Specifically, a sequential stochastic strategy based on *Gillespie's theory of stochastic kinetics* [9] was introduced in order to overcome the two

previous problems when developing a modeling framework for cellular systems biology based on P systems [16]. Here we refer to this variant as *stochastic P systems.*

Definition 1 (Stochastic P Systems). *A Stochastic P system is a construct:*

$$\Pi = ((\Sigma_{obj}, \Sigma_{str}), L, \mu, M_{l_1}, \ldots, M_{l_m}, (R_{l_1}^{obj}, R_{l_1}^{str}), \ldots, (R_{l_m}^{obj}, R_{l_m}^{str})),$$

where:

- Σ_{obj} *is a finite alphabet of objects representing molecular species whose internal structure is not relevant in the functioning of the system under study.*
- Σ_{str} *is a finite alphabet of objects representing relevant parts of some molecular species in the system. These objects are arranged into strings describing the structure of molecular species.*
- $L = \{l_1, \ldots, l_m\}$ *is a finite alphabet of symbols representing compartment labels used to identify compartment classes. Compartments with the same label share the same class, i.e., set of rewriting rules and initial multisets.*
- μ *is a membrane structure consisting of* $n \geq 1$ *membranes defining compartments identified in a one to one manner with values from* $\{1, \ldots, n\}$ *and labeled with elements from* L*.*
- $M_{l_t} = (w_t, s_t)$*, for each* $1 \leq t \leq m$*, is the initial state of the compartments from the class identified by label* l_t*, where* $w_t \in \Sigma_{obj}^*$ *is a finite multiset of individual objects and* s_t *is a finite set of strings over* Σ_{str}*. A multiset of objects, obj is represented as* $obj = o_1 + o_2 + \ldots + o_p$ *with* $o_1, \ldots, o_p \in \Sigma_{obj}$*. Strings are represented as follows* $\langle s_1 \cdot s_2 \cdots s_q \rangle$ *where* $s_1, \ldots, s_q \in \Sigma_{str}$*.*
- $R_{l_t}^{obj} = \{r_1^{obj,l_t}, \ldots, r_{k_{obj,l_t}}^{obj,l_t}\}$*, for each* $1 \leq t \leq m$*, is a finite multiset of rewriting rules on multisets of objects associated with compartments of the type specified by the label* l_t*. The rewriting rules on multisets of objects are of the following form:*

$$r_j^{obj,l_t} : \ obj_1 \, [\, obj_2 \,]_l \overset{c_j^{obj,l_t}}{\longrightarrow} obj_1' \, [\, obj_2' \,]_l \tag{1}$$

with $obj_1, obj_2, obj_1', obj_2'$ *some finite multisets of objects from* Σ_{obj} *and* l *a label from* L*. These rules are multiset rewriting rules that operate on both sides of membranes, that is, a multiset* obj_1 *placed outside a membrane labeled by* l *and a multiset* obj_2 *placed inside the same membrane can be simultaneously replaced with a multiset* obj_1' *and a multiset* obj_2'*, respectively.*

Note that a constant c_j^{obj,l_t} *is associated specifically with each rule. This constant will be referred to as stochastic constant and is key to provide P systems with a stochastic extension as it will be used to compute the probability and time needed to apply each rule. This constant depends only on the physical properties of the molecules and compartments involved in the reaction described by the rule like temperature, pressure, pH, volume, etc.*
- $R_{l_t}^{str} = \{r_1^{str,l_t}, \ldots, r_{k_{str,l_t}}^{str,l_t}\}$*, for each* $1 \leq t \leq m$*, is a finite set of rewriting rules on multisets of strings and objects associated with compartments of the type defined by* l_t *and of the following form:*

$$r_j^{str,l_t} : \ [\, obj + str \,]_l \overset{c_j^{str,l_t}}{\longrightarrow} [\, obj' + str'; str_1' + \ldots + str_s' \,]_l \tag{2}$$

with obj, obj' multisets of objects over Σ_{obj} and $str, str', str'_1, \ldots, str'_s$ strings over Σ_{str}. These rules operate on both multisets of objects and strings. The objects obj are replaced by the objects obj'. Simultaneously a substring str is replaced by str' whereas the strings $str'_1, \ldots, str'_s$ are produced to form part of the content of the compartment. In the same way as for rewriting rules on multisets of objects a stochastic constant c_j^{str,l_t} is associated with each rule.

The previous definition is provided with a stochastic strategy for the application of the rewriting rules by extending the Gillespie algorithm to the multicompartmental structure of P systems. The resulting algorithm has been referred to as the *Multicompartmental Gillespie Algorithm* (MGA) [16]. The Gillespie algorithm [9] can only be applied directly in a single, fixed and well mixed volume. In our approach the first step consists of treating each compartment defined by a membrane as a fixed and well mixed volume where the rewriting rule to be applied and the elapsed time before its application is computed using the Gillespie Direct Method. Our algorithm then applies the corresponding rules following the order determined by these waiting times. After the application of each rule the algorithm recomputes the rules to be applied and the waiting times in the compartments affected by the application of the last rule using the Gillespie Direct Method. Finally, the MGA halts when a prefixed simulation time is reached or no further rules can be applied.

3 Running Example

In order to illustrate our modeling framework we will use an abstract gene regulation system inspired from the functioning and structure of the *lac operon* in *Escherichia coli* (*E. coli*). This operon consists of three structural genes, *lacZ*, *lacY* and *lacA*, located sequentially on the genome and transcribed into one single mRNA. Their protein products are involved in the sensing, uptake and metabolism of lactose. The transcription of the *lac operon* is both positively and negatively regulated and it is considered a canonical example of gene transcription regulation in prokaryotes [18].

The linear structure of the *lac operon* (Figure 1) starts with a region called *cap* where the activator protein CRP binds and increases the rate of transcription. Following this site there is an operator sequence that we will refer to as *op* where the repressor protein LacI binds to stop transcription. The structural genes *lacZ*, *lacY* and *lacA* then follow. The first gene *lacZ* codifies the enzyme β-galactosidase involved in the metabolism of lactose by cleaving it into glucose and galactose; allolactose appears as a byproduct of this reaction. The protein product of the second gene *lacY* is a permease that associates to the cell membrane and acts as a pump transporting lactose into the cell. The function of the protein coded in the last gene *lacA* is not yet fully understood.

The regulation of the *lac operon* allows *E. coli* to express the genes in the operon only when it is more beneficial for the cell. In the absence of lactose in the media the repressor LacI binds to the operator *op* preventing the structural

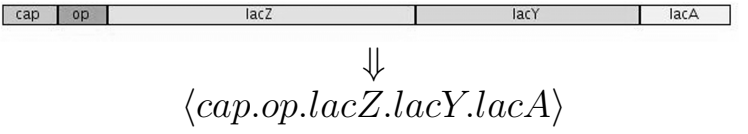

Fig. 1. A schematic representation of the structure of the *lactose operon* (top) and its representation as a string (bottom)

genes from being transcribed since they are not needed under these conditions. Nevertheless, occasionally the repressor drops from the operator producing a basal transcription of the operon.

When lactose becomes available it starts to be transported inside the cell by the basal number of LacY proteins on the cell surface. Once in the cytoplasm it interacts with the basal number of β-galactosidase producing as a byproduct *allolactose*. Allolactose in turn binds to the repressor LacI and changes its state so it cannot bind to the operator allowing transcription of the structural genes. The resulting increase in production of LacY and β-galactosidase forms a positive feed-back loop increasing the number of allolactose molecules which interact with the repressors preventing premature termination of transcription.

The *lac operon* is also under positive regulation by the protein CRP. This protein is activated by the glucose transport system and when active it binds to the *cap* site facilitating transcription. Even in the presence of lactose if glucose is present in the media CRP will not be active as the transport system will be occupied, pumping glucose into the cell. Therefore CRP will not bind to the operon to assist transcription. Only in the presence of lactose and absence of glucose will CRP be active and bound to the operon, producing the full transcription of the operon.

This gene regulation system will be used in the following section as the running example illustrating our modeling principles.

4 Modeling Principles

The complexity of cellular systems is organized into different levels ranging from the molecular to the cellular and colony scales. These levels of complexity are not independent instead they are tightly interrelated influencing each other directly. In this respect, stochastic P systems present an integrating multiscale modeling framework which explicitly specifies the molecular, cellular and colony levels in cellular systems in a relevant and understandable manner.

One of our research lines consists of the development of integrative modeling principles within the modeling framework of stochastic P systems. More specifically we will present some ideas on how to describe molecular species, cellular regions and compartments, molecular interactions, gene expression control and cell colonies. Our running example will be used to illustrate our modeling principles.

- **Molecular species:** These are specified as individual objects or strings of objects. Molecules with an internal structure that is relevant in the

functioning of the system are specified using strings. For example, gene operons with a linear structure consisting of promoters, operators, transcription/translation starting points, etc, otherwise molecular species are described using individual objects.

Table 1. Specification of the molecular species in the *lac operon*

Molecular Species	**Object**
RNA Polymerase	$RNAP$
Ribosome	Rib
Repressor	$LacI$
Activator	$CRP\ \ CRP^*$
LacZ product	$LacZ$
LacY product	$LacY$
LacA product	$LacA$
Lactose	$Lact$
Allolactose	$Allolac$
Glucose	$Gluc$
Glucose transport system	$Gluc$
Complex glucose transport system	$Gluc\text{-}GTS$
Complex lactose LacY product	$Lact\text{-}LacY$
Complex lactose LacZ product	$Lact\text{-}LacZ$
Complex lactose LacZ product	$Lact\text{-}LacZ$
Complex allolactose repressor	$Allolac\text{-}LacI$

Operon site	**Object**
Activator binding site	cap
Occupied activator binding site	cap^{CRP^*}
Repressor binding site	op
Occupied repressor binding site	op^{LacI}
lacZ gene	$lacZ$
lacY gene	$lacY$
lacA gene	$lacA$
lacZ mRNA	$mlacZ$
lacY mRNA	$mlacY$
lacA mRNA	$mlacA$

Running example: The different molecular species in our example will be specified according to this modeling principle. On the one hand, the proteins and complexes of proteins involved in the regulation and expression of the *lac operon* are specified as individual objects since we are not interested in their internal structure (Table 1). On the other hand, each component of the *lac operon* will be described using an object such that the *lac operon* structure is specified as a string containing these objects in the specific order they can be found in E. coli's genome (Figure 1).

- **Membranes:** Compartmentalization and membranes are fundamental in the structural organization and functioning of living cells. Membranes do not act as passive boundaries of cells and compartments; instead they play a key role in the regulation of the metabolism and information processing between the outside and the inside of compartments. P systems constitutes one of the few computational frameworks which explicitly specifies compartments and membranes. For instance, P systems have been used to study selective

uptake of molecules from the environment [20], signalling at the cell surface [12] and colonies of interacting bacteria which communicate by sending and receiving diffusing signals [2,21]. In general P system membranes are used to define relevant regions in cellular systems and therefore they do not always correspond to real cell membranes although normally they do.

Running example: In the *lac operon* gene regulation system there are two relevant regions. Namely, the bacterium surface where $LacY$ and GTS act as pumps transporting lactose and glucose into the cell and the aqueous interior or cytoplasm where the operon is located together with the different transcription factors and proteins. These two regions are represented using two membranes embedded one inside the other to described the structure of an E. coli bacterium (Figure 2).

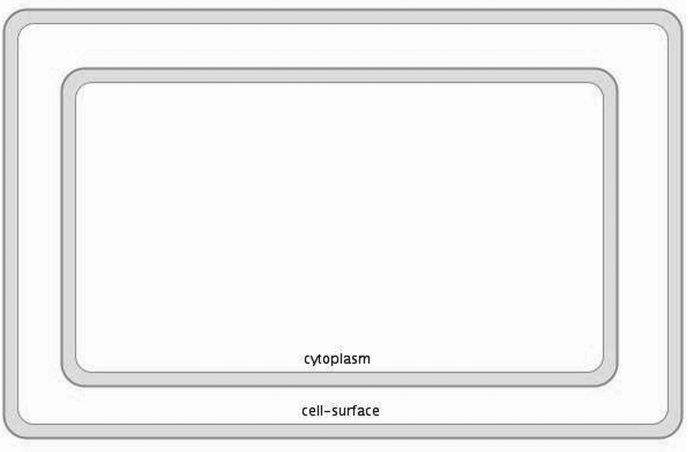

Fig. 2. Graphical representation of the membrane structure specifying an E. coli bacterium

- **Molecular processes consisting of protein-protein interactions and protein translocation:** Such processes are normally described in P systems using rewriting rules on multisets of objects. Our P system modeling framework aims at providing a comprehensive and relevant rule-based schema for the most common molecular interactions taking place in living cells. More specifically, our approach focuses on the transformation and degradation of molecular species, the formation and dissociation of complexes, and the basic processes of communication and transport between different compartments in cellular systems (Table 2).

  Running example: The protein-protein interactions in our gene regulation system are described using the rewriting rules on multisets of objects presented in Table 3. Rules r_{29}, r_{30}, r_{31} and r_{32} are examples of complex formation and dissociation rules. The degradation and dilution of different proteins

Table 2. P system rule-based schemas for the most common molecular interactions

Molecular Interaction	**P System Rules**
Transformation and Degradation	$[\ a\ ]_l \xrightarrow{c} [\ b\ ]_l \qquad [\ a\ ]_l \xrightarrow{c} [\ ]_l$
Complex formation and dissociation	$[\ a+b\ ]_l \xrightarrow{c_f} [\ c\ ]_l \quad [\ c\ ]_l \xrightarrow{c_d} [\ a+b\ ]_l$
Diffusion in and out	$a\ [\ ]_l \xrightarrow{c_{in}} [\ a\ ]_l \qquad [\ a\ ]_l \xrightarrow{c_{out}} a\ [\ ]_l$
Binding and debinding	$a\ [\ b\ ]_l \xrightarrow{c_{lb}} [\ c\ ]_l \qquad [\ c\ ]_l \xrightarrow{c_{ld}} a\ [\ b\ ]_l$
Recruitment and releasing	$a\ [\ b\ ]_l \xrightarrow{c_{rt}} c\ [\ ]_l \quad c\ [\ ]_l \xrightarrow{c_{rl}} a\ [\ b\ ]_l$

is specified in rules r_{22}, r_{23} and r_{24}. Finally, active uptake of glucose and lactose are modeled using the binding and releasing rules r_{27}, r_{28}, r_{33} and r_{34}.

- **Gene expression control:** The sensing of signals and the processing of the information they convey is performed in living cells through molecular interactions of the type presented in Table 2. The response of cells to these signals consists of the expression of appropriate proteins codified in specific genes. Gene expression control has been described in P systems using either rewriting rules on multisets of objects or rewriting rules on multisets of objects and strings according to the structural organization of the genes in the system under study. Tables 4 and 5 presents these two alternatives for the specification of the most important processes in gene expression control; transcription factor binding and debinding, transcription and translation.

  From a simplistic point of view the processes involved in transcription factor binding and debinding, transcription and translation can be represented by individual rewriting rules on multisets of objects (Table 4). Nevertheless, these processes are very complex and they consist of different stages like operator/promoter recognition by transcription factors and RNA polymerase, transcription/translation initiation/termination, elongation, etc. A more accurate and detailed description of all these processes is achieved by using rewriting rules on multisets of strings and objects of the form of the rules in Table 5.

  Running Example: The gene regulation control in the *lac operon* is modeled using the rewriting rules on multisets of objects and strings given in Table 3. More specifically, the binding and debinding of the activator and repressor to their corresponding binding sites is represented using rules r_3, r_4, r_7 and r_8. Transcription initiation in the presence and absence of the promoter site occupied by the activator CRP^* is specified using rules r_1, r_2, r_5 and r_6. The transcription of the structural genes *lacZ*, *lacY* and *lacA* is described by the rules r_9, r_{10}, r_{11} and r_{12}. Finally, translation and mRNA degradation is modeled with the rules r_{13} - r_{21}.
- **Cell colonies:** The last level of organization that has been represented using P systems consists of cellular systems where cells form colonies by interacting

Table 3. Lac Operon Regulation Rules

Nr.	Rule	Stochastic Constant
r_1 :	$[\ RNAP + \langle cap \rangle\ ]_b \xrightarrow{c_1} [\ \langle cap.RNAP \rangle\ ]_b$	$c_1 = 5 \times 10^{-3} min^{-1}$
r_2 :	$[\ \langle cap.RNAP \rangle\ ]_b \xrightarrow{c_2} [\ RNAP + \langle cap \rangle\ ]_b$	$c_2 = 1 min^{-1}$
r_3 :	$[\ CRP^* + \langle cap \rangle\ ]_b \xrightarrow{c_3} [\ \langle cap^{CRP^*} \rangle\ ]_b$	$c_3 = 16.6 min^{-1}$
r_4 :	$[\ \langle cap^{CRP^*} \rangle\ ]_b \xrightarrow{c_4} [\ CRP^* + \langle cap \rangle\ ]_b$	$c_4 = 10 min^{-1}$
r_5 :	$[\ RNAP + \langle cap^{CRP^*} \rangle\ ]_b \xrightarrow{c_5} [\ \langle cap^{CRP^*}.RNAP \rangle\ ]_b$	$c_5 = 0.2 min^{-1}$
r_6 :	$[\ \langle cap^{CRP^*}.RNAP \rangle\ ]_b \xrightarrow{c_6} [\ RNAP + \langle cap^{CRP^*} \rangle\ ]_b$	$c_6 = 1 min^{-1}$
r_7 :	$[\ LacI + \langle op \rangle\ ]_b \xrightarrow{c_7} [\ \langle op^{LacI} \rangle\ ]_b$	$c_7 = 166 min^{-1}$
r_8 :	$[\ \langle op^{LacI} \rangle\ ]_b \xrightarrow{c_8} [\ LacI + \langle op \rangle\ ]_b$	$c_8 = 0.1 min^{-1}$
r_9 :	$[\ \langle RNAP.op \rangle\ ]_b \xrightarrow{c_9} [\ \langle op.RNAP \rangle\ ]_b$	$c_9 = 3 min^{-1}$
r_{10} :	$[\ \langle RNAP.lacZ \rangle\ ]_b \xrightarrow{c_{10}} [\ \langle lacZ.RNAP \rangle ; \langle mlacZ \rangle\ ]_b$	$c_{10} = 0.78 min^{-1}$
r_{11} :	$[\ \langle RNAP.lacY \rangle\ ]_b \xrightarrow{c_{11}} [\ \langle lacY.RNAP \rangle ; \langle mlacY \rangle\ ]_b$	$c_{11} = 1.92 min^{-1}$
r_{12} :	$[\ \langle RNAP.lacA \rangle\ ]_b \xrightarrow{c_{12}} [\ RNAP + \langle lacA \rangle ; \langle mlacA \rangle\ ]_b$	$c_{12} = 4 min^{-1}$
r_{13} :	$[\ Rib + \langle mlacZ \rangle\ ]_b \xrightarrow{c_{13}} [\ \langle Rib.mlacZ \rangle\ ]_b$	$c_{13} = 0.12 min^{-1}$
r_{14} :	$[\ Rib + \langle mlacY \rangle\ ]_b \xrightarrow{c_{14}} [\ \langle Rib.mlacY \rangle\ ]_b$	$c_{14} = 0.12 min^{-1}$
r_{15} :	$[\ Rib + \langle mlacA \rangle\ ]_b \xrightarrow{c_{15}} [\ \langle Rib.mlacA \rangle\ ]_b$	$c_{15} = 0.12 min^{-1}$
r_{16} :	$[\ \langle Rib.mlacZ \rangle\ ]_b \xrightarrow{c_{16}} [\ Rib + LacZ + \langle mlacZ \rangle\ ]_b$	$c_{16} = 0.12 min^{-1}$
r_{17} :	$[\ \langle Rib.mlacY \rangle\ ]_b \xrightarrow{c_{17}} [\ Rib + LacY + \langle mlacY \rangle\ ]_b$	$c_{17} = 1.73 min^{-1}$
r_{18} :	$[\ \langle Rib.mlacA \rangle\ ]_b \xrightarrow{c_{18}} [\ Rib + LacA + \langle mlacA \rangle\ ]_b$	$c_{18} = 3.55 min^{-1}$
r_{19} :	$[\ \langle mlacZ \rangle\ ]_b \xrightarrow{c_{19}} [\ ]_b$	$c_{19} = 6 \times 10^{-3} min^{-1}$
r_{20} :	$[\ \langle mlacY \rangle\ ]_b \xrightarrow{c_{20}} [\ ]_b$	$c_{20} = 6 \times 10^{-3} min^{-1}$
r_{21} :	$[\ \langle mlacA \rangle\ ]_b \xrightarrow{c_{21}} [\ ]_b$	$c_{21} = 6 \times 10^{-3} min^{-1}$
r_{22} :	$[\ LacZ\ ]_b \xrightarrow{c_{22}} [\ ]_b$	$c_{22} = 6.9 \times 10^{-2} min^{-1}$
r_{23} :	$[\ LacY\ ]_b \xrightarrow{c_{23}} [\ ]_b$	$c_{23} = 6.9 \times 10^{-2} min^{-1}$
r_{24} :	$[\ LacA\ ]_b \xrightarrow{c_{24}} [\ ]_b$	$c_{24} = 6.9 \times 10^{-2} min^{-1}$
r_{25} :	$[\ LacY\ ]_b \xrightarrow{c_{25}} LacY\ [\ ]_b$	$c_{25} = 1 min^{-1}$
r_{26} :	$LacY\ [\ ]_b \xrightarrow{c_{26}} [\ LacY\ ]_b$	$c_{26} = 0.7 min^{-1}$
r_{27} :	$Lact\ [\ LacY\ ]_s \xrightarrow{c_{27}} [\ Lact\text{-}LacY\ ]_s$	$c_{27} = 10 min^{-1}$
r_{28} :	$Lact\text{-}LacY\ [\ ]_b \xrightarrow{c_{28}} LacY\ [\ Lact\ ]_b$	$c_{28} = 10 min^{-1}$
r_{29} :	$[\ Lact + LacZ\ ]_b \xrightarrow{c_{29}} [\ Lact\text{-}LacZ\ ]_b$	$c_{29} = 10 min^{-1}$
r_{30} :	$[\ Lact\text{-}LacZ\ ]_b \xrightarrow{c_{30}} [\ Allolac + LacZ\ ]_b$	$c_{30} = 10 min^{-1}$
r_{31} :	$[\ Allolac + LacI\ ]_b \xrightarrow{c_{31}} [\ Allolac\text{-}LacI\ ]_b$	$c_{31} = 1 min^{-1}$
r_{32} :	$[\ Allolac\text{-}LacI\ ]_b \xrightarrow{c_{32}} [\ Allolac + LacI\ ]_b$	$c_{32} = 10^{-4} min^{-1}$
r_{33} :	$Gluc\ [\ GTS\ ]_s \xrightarrow{c_{33}} [\ Gluc\text{-}GTS\ ]_s$	$c_{33} = 1 min^{-1}$
r_{34} :	$Gluc\text{-}GTS\ [\ ]_b \xrightarrow{c_{34}} GTS\ [\ Gluc\ ]_b$	$c_{34} = 10 min^{-1}$
r_{35} :	$GTS\ [\ CRP\ ]_b \xrightarrow{c_6} GTS\ [\ CRP^*\ ]_b$	$c_{35} = 6.9 \times 10^{-3} min^{-1}$
r_{36} :	$[\ CRP^*\ ]_b \xrightarrow{c_6} [\ ]_b$	$c_{36} = 0.069 min^{-1}$

and exhibiting coordinated behavior. The specification of the environment where the colony of cell is located cannot always be described by a single membrane since it is normally too big to be considered a well mixed volume or region where the Gillespie Algorithm can be applied. In this respect,

Table 4. P system rule-based schemas for gene expression control using multisets of objects

Molecular Interaction	Rules on Multisets of Objects
Transcription Factor Binding and Debinding	$[\, Tf + gene \,]_l \xrightarrow{c_{on}} [\, Tf\text{–}gene \,]_l$ $[\, Tf\text{–}gene \,]_l \xrightarrow{c_{off}} [\, Tf + gene \,]_l$
Transcription	$[\, gene \,]_l \xrightarrow{c_{tc}} [\, gene + rna \,]_l$
Translation	$[\, rna \,]_l \xrightarrow{c_{tl}} [\, rna + prot \,]_l$

Table 5. P system rule-based schemas for gene expression control using multisets of objects and strings

Molecular Interaction	Rules on Multisets of Strings and Objects
Transcription Factor Binding and Debinding	$[\, Tf + \langle op \rangle \,]_l \xrightarrow{c_{on}} [\, \langle op' \rangle \,]_l$ $[\, \langle op' \rangle \,]_l \xrightarrow{c_{off}} [\, Tf + \langle op \rangle \,]_l$
Transcription	$[\, \mathrm{RNAP} + \langle prom \rangle \,]_l \xrightarrow{c_{rb}} [\, \langle prom.\mathrm{RNAP} \rangle \,]_l$ $[\, \langle \overline{s}_0.w.\mathrm{RNAP}.s_N \rangle \,]_l \xrightarrow{c_{el}} [\, \langle s_N.\overline{s}_0.w.\overline{s}_N.\mathrm{RNAP} \rangle \,]_l$ $[\, \langle \overline{s}_0.w.\mathrm{RNAP}.s_t \rangle \,]_l \xrightarrow{c_{ter}} [\, \mathrm{RNAP} + \langle s_t \rangle;\ \langle \overline{s}_0.w.\overline{s}_t \rangle]_l$
Translation	$[\, \mathrm{Rib} + \langle \overline{s}_0 \rangle \,]_l \xrightarrow{c_{tli}} [\, \langle \overline{site}_0.\mathrm{Rib} \rangle \,]_l$ $[\, \langle \mathrm{Rib}.\overline{s}_N \rangle \,]_l \xrightarrow{c_{tle}} [\, \langle \overline{s}_N.\mathrm{Rib} \rangle \,]_l$ $[\, \langle \mathrm{Rib}.\overline{s}_t \rangle \,]_l \xrightarrow{c_{tlt}} [\, \mathrm{Rib} + \mathrm{Prot} + \langle \overline{s}_t \rangle \,]_l$

the environment is divided into a set of small regions that can be considered well mixed volumes. These regions are then connected according to a graph defining the topology of the environment. This structure has been termed multienvironment [11]. A cell colony is then specified as a collection of individual P systems representing individual cells distributed over the multienvironment. These P systems interact by passive or active transport rules using some of the specifications of molecular interaction described previously. The different regions in a multienvironment can also interact by passive diffusion rules of the following form:

$$[\, obj \,]_l - [\ ]_{l'} \xrightarrow{c} [\ ]_l - [\, obj \,]_{l'} \tag{3}$$

These rules are multiset rewriting rules that operate on two environments, one labeled l which is linked to another environment labeled l'. A multiset obj is removed from the first environment and placed in the second one. In this way, we are able to capture in a concise way the diffusion of signals from one region to another in a large environment. As well as objects, P systems

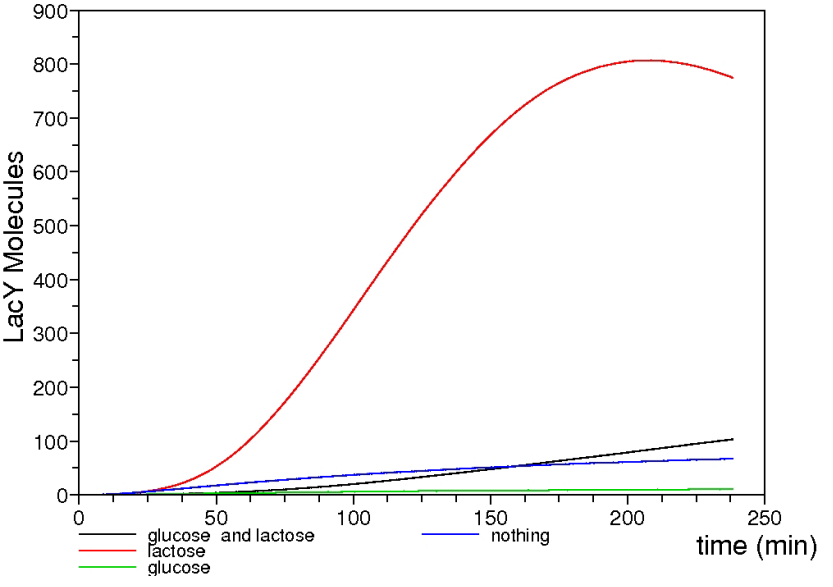

Fig. 3. Evolution over time of the average of the LacY protein products over a colony of 1000 bacteria for four different environmental conditions representing the presence and/or absence of glucose and/or lactose

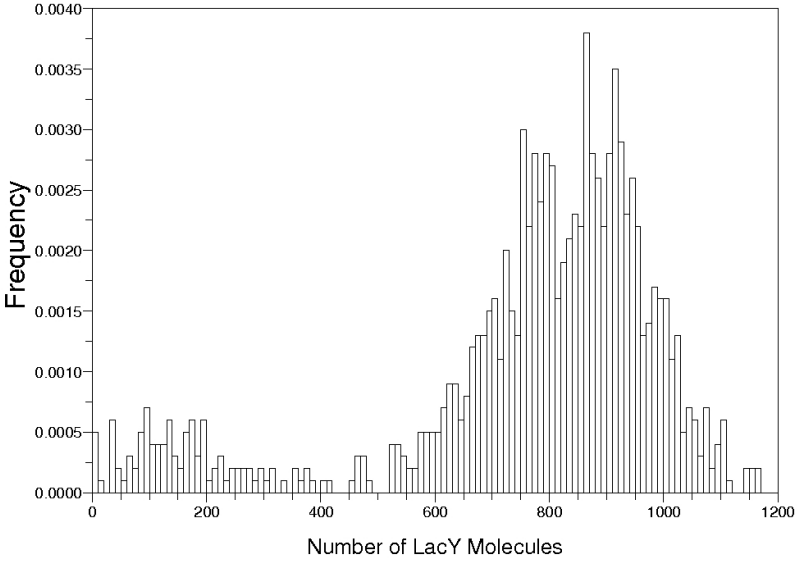

Fig. 4. Histogram representing the frequency of the number of LacY proteins over a colony of 1000 bacteria

representing individual cells can be moved from one environment to another using the following type of rules:

$$[\,[\,]_{l''}\,]_{l} - [\ ]_{l'} \xrightarrow{c} [\ ]_{l} - [\,[\,]_{l''}\,]_{l'} \tag{4}$$

When a rule of this type is applied, a membrane with label l'' and all its contents, objects and other membranes, are moved from an environment labeled l to another connected to it that must be labeled l'. This colony level specification of P systems was introduced in [2] and was used to model the quorum sensing system in the marine bacterium *Vibrio fischeri* from an artificial life perspective in [21].

Running Example: In our case study we have examined the behavior of a colony of 1000 bacteria located inside a single environment. Each bacterium was represented using the membrane structure in Figure 2, the rewriting rules in Table 3, the objects in Table 1 and the string in Figure 1. We have run simulations for four different environmental conditions representing the presence and/or absence of glucose and/or lactose.

We computed the average across the entire colony of the number of LacY protein products over time (Figure 3). Our results were in agreement with the known behavior of *E. coli* which prefers to consume glucose to lactose. Only in the absence of glucose and presence of lactose will *E. coli* fully express the *lac operon* and activate the processes involved in the uptake and metabolism of lactose.

In order to get a more detailed idea of the behavior of the colony we studied the frequency of the number of LacY protein products over all the bacteria in the case when only lactose is present in the environment (Figure 4). The number of LacY protein products over the colony is a bimodal distribution. Most bacteria fully express the *lac operon*, whereas a small but still noticeable fraction of the colony does not activate the uptake and metabolism of lactose. This type of behavior is characteristic of gene regulation systems with positive feedback loops [1].

5 Modularization

Cellular processes arise as emergent behavior from the interactions between many molecular species. It has been postulated that although these interactions are apparently messy they are arranged in a modular way. Our P system modeling framework supports the use of modules of rules to represent biological functions that are separable or orthogonal to some extend from the functioning of the rest of the system. A P system module is a set of rewriting rules of the forms previously introduced describing molecular processes occurring frequently in cell systems. A module is identified with a name and three sets of variables, V, representing the molecular species; C, the stochastic constants associated with each rules; and Lab, the labels of the compartments involved in the rules. Since a module is a set of rules starting from simple modules more complex modules

can be constructed by applying set union. Table 6 presents example P system modules describing the most common regulation mechanism in gene expression.

Following this idea one of our current research lines focuses on the automatic development of cellular models exhibiting a prefixed behavior by assembling P system modules automatically. Our methodology optimizes both the modular structure of the P systems and the stochastic constants associated with the rules. Specifically, our methodology consists of two nested genetic algorithms: the first one evolves the combination of modules or modular structure of the model whereas the second one optimizes the stochastic constants associated with the different rules in the modules. Our approach is incremental, by starting with simple predefined modules from an elementary library newly generated modules obtained by combining these elementary modules are added to the library after having been analyzed and validated. This allows us to develop more intricate and circuitous modular structures. Our methodology has been tested on three case studies, namely, molecular complexation, enzymatic reactions and autoregulation in transcriptional networks [22].

Table 6. Three examples of P system modules describing the most common regulation mechanisms in gene expression

Molecular process	**P System Module**
Constitutive Expression	$[\ gene\ ]_l \xrightarrow{c_1} [\ gene + rna\ ]_l$ $[\ rna\ ]_l \xrightarrow{c_2} [\ ]_l$ $[\ rna\ ]_l \xrightarrow{c_3} [\ rna + p\ ]_l$ $[\ p\ ]_l \xrightarrow{c_4} [\ ]_l$
Positive Regulation	$[\ a + gene\ ]_l \xrightarrow{c_1} [\ a.gene\ ]_l$ $[\ a.gene\ ]_l \xrightarrow{c_2} [\ a + gene\ ]_l$ $[\ a.gene\ ]_l \xrightarrow{c_3} [\ a.gene + rna\ ]_l$ $[\ rna\ ]_l \xrightarrow{c_4} [\ ]_l$ $[\ rna\ ]_l \xrightarrow{c_5} [\ rna + p\ ]_l$ $[\ p\ ]_l \xrightarrow{c_6} [\ ]_l$
Negative Regulation	$[\ gene\ ]_l \xrightarrow{c_1} [\ gene + rna\ ]_l$ $[\ r + gene\ ]_l \xrightarrow{c_2} [\ r.gene\ ]_l$ $[\ r.gene\ ]_l \xrightarrow{c_3} [\ r + gene\ ]_l$ $[\ rna\ ]_l \xrightarrow{c_4} [\ ]_l$ $[\ rna\ ]_l \xrightarrow{c_5} [\ rna + p\ ]_l$ $[\ p\ ]_l \xrightarrow{c_6} [\ ]_l$

In cellular systems, modularization does not only arise from chemical specificity, but is also determined by spatial localization of molecular species in different compartments. The P system modules introduced so far only describe modularity due to chemical specificity. No geometric information is associated with any components of P systems. Recently we have proposed to extend population P systems by using finite lattices on which individual P systems are distributed. These P systems communicate by sending and receiving objects according to rules of the following form:

$$[\ obj\ ]_l \overset{\mathbf{v}}{\Leftrightarrow} [\ \ ]_{l'} \xrightarrow{c} [\ ]_l \overset{\mathbf{v}}{\Leftrightarrow} [\ obj\ ]_{l'} \tag{5}$$

The application of a rule of the previous form in a P system Π_j located in position $\mathbf{p}$, moves the objects obj from the skin membrane l of Π_j to the skin membrane l' of P system $\Pi_{j'}$ located in position $\mathbf{p} + \mathbf{v}$. The stochastic constant c associated with the rule plays the same role as in the previous cases.

6 Conclusions and Future Work

In this paper we have presented P systems as a modeling approach within *executable biology* able to specify and simulate multiscale systems ranging from the molecular to the cell and colony level. The modeling principles used in P systems for the specification of molecular species, networks of interacting molecules, individual cells and collections of cells have been discussed. A running example consisting in an abstract gene regulation system based on the *lac operon* has been used to illustrate our approach. Our results show the characteristic bimodal behavior of a colony of bacterial cells with a positive feed back loop.

This framework is currently being extended and implemented in a software system with integrates stochastic simulation of multicellular P system models with analytic techniques based on model checking and automate model generation through the assembling of P systems modules. This framework is also being used to develop models of plant hormone transport in *Arabidopsis thaliana* and quorum sensing in *Pseudomonas aeruginosa*.

Acknowledgements. We would like to acknowledge EPSRC grant EP/E017215/1 and BBSRC grants BB/F01855X/1 and BB/D019613/1.

References

1. Alon, U.: Network motifs: theory and experimental approaches. Nature Reviews Genetics 8, 450–461 (2007)
2. Bernardini, F., Gheorghe, M., Krasnogor, N.: Quorum sensing P systems. Theoretical Computer Sci. 371, 20–33 (2007)
3. Besozzi, D., Cazzaniga, P., Pescini, D., Mauri, G., Colombo, S., Martegani, E.: Modeling and stochastic simulation of the Ras/cAMP/PKA pathway in the yeast Saccharomyces cerevisiae evidences a key regulatory function for intracellular guanine nucleotides pools. Journal of Biotechnology 133, 377–385 (2008)
4. Bianco, L., Fontana, F., Manca, V.: P systems with reaction maps. Intern. J. Foundations of Computer Sci. 17, 27–48 (2006)
5. Ciobanu, G., Pan, L., Păun, G., Pérez-Jiménez, M.J.: P systems with minimal parallelism. Theoretical Computer Sci. 378, 117–130 (2007)
6. Fisher, J., Henzinger, T.A.: Executable cell biology. Nature Biotechnology 25, 1239–1249 (2007)
7. Fontana, F., Manca, V.: Discrete solutions to differential equations by metabolic P systems. Theoretical Computer Sci. 372, 165–182 (2007)
8. Freund, R.: P systems working in the sequential mode on arrays and strings. Int. J. Found. Comput. Sci. 16, 663–682 (2005)
9. Gillespie, D.T.: Stochastic simulation of chemical kinetics. Annu. Rev. Phys. Chem. 58, 35–55 (2007)
10. Heiner, M., Gilbert, D., Donaldson, R.: Petri nets for systems and synthetic biology. In: Bernardo, M., Degano, P., Zavattaro, G. (eds.) SFM 2008. LNCS, vol. 5016, pp. 215–264. Springer, Heidelberg (2008)
11. Krasnogor, N., Gheorghe, M., Terrazas, G., Diggle, S., Williams, P., Camara, M.: An appealing computational mechanism drawn from bacterial quorum sensing. Bulletin of the EATCS 85, 135–148 (2005)

12. Păun, A., Jesús Pérez-Jímenez, M., Romero-Campero, F.J.: Modeling signal transduction using P systems. In: Hoogeboom, H.J., Păun, G., Rozenberg, G., Salomaa, A. (eds.) WMC 2006. LNCS, vol. 4361, pp. 100–122. Springer, Heidelberg (2006)
13. Păun, G.: Computing with membranes. J. Computer and System Sci. 61, 108–143 (2000)
14. Păun, G.: Membrane Computing: An Introduction. Springer, Heidelberg (2002)
15. Păun, G., Romero-Campero, F.J.: Membrane computing as a modeling framework. Cellular systems case studies. In: Bernardo, M., Degano, P., Zavattaro, G. (eds.) SFM 2008. LNCS, vol. 5016, pp. 168–214. Springer, Heidelberg (2008)
16. Jesús Pérez-Jímenez, M., Romero-Campero, F.J.: P systems, a new computational modelling tool for systems biology. In: Priami, C., Plotkin, G. (eds.) Transactions on Computational Systems Biology VI. LNCS (LNBI), vol. 4220, pp. 176–197. Springer, Heidelberg (2006)
17. Pescini, D., Besozzi, D., Mauri, G., Zandron, C.: Dynamical probabilistic P systems. Intern. J. Foundations of Computer Sci. 17, 183–204 (2006)
18. Ptashne, M., Gann, A.: Genes and Signals. Cold Spring Harbor Laboratory Press (2002)
19. Regev, A., Shapiro, E.: The π-calculus as an abstraction for biomolecular systems. Modelling in Molecular Biology, 1–50 (2004)
20. Romero-Campero, F.J., Pérez-Jiménez, M.J.: Modelling gene expression control using P systems: the Lac Operon, a case study. BioSystems 91, 438–457 (2008)
21. Romero-Campero, F.J., Pérez-Jiménez, M.J.: A model of the quorum sensing system in Vibrio fischeri using P systems. Artificial Life 14, 1–15 (2008)
22. Romero-Campero, F.J., Cao, H., Cámara, M., Krasnogor, N.: Structure and parameter estimation for cell systems biology models. In: Proc. of the Genetic and Evolutionary Computation Conference, Atlanta, USA, pp. 331–338 (2008)
23. Romero-Campero, F.J., Twycross, J., Cámara, M., Bennett, M., Gheorghe, M., Krasnogor, N.: Modular assembly of cell systems biology models using P systems (submitted)