An Approach to the Engineering of Cellular Models Based on P Systems

Francisco J. Romero-Campero and Natalio Krasnogor

Automated Scheduling, Optimisation and Planning Research Group
School of Computer Science, Jubilee Campus, University of Nottingham
Nottingham NG8 1BB, United Kingdom
{fxc,nxk}@cs.nott.ac.uk

Extended Abstract

Living cells assembled into colonies or tissues communicate using *complex systems*. These systems consist in the interaction between many molecular species distributed over many compartments. Among the different cellular processes used by cells to monitor their environment and respond accordingly, gene regulatory networks, rather than individual genes, are responsible for the information processing and orchestration of the appropriate response [16].

In this respect, *synthetic biology* has emerged recently as a novel discipline aiming at unravelling the design principles in gene regulatory systems by synthetically engineering transcriptional networks which perform a specific and prefixed task [2]. Formal modelling and analysis are key methodologies used in the field to engineer, assess and compare different genetic designs or devices.

In order to model cellular systems in colonies or tissues one requires a formalism able to represent the following relevant features:

- Single cells should be described as the *elementary units* in the system. Nevertheless, they cannot be represented as homogeneous points as they exhibit *complex structures* containing different compartments where specific molecular species interact according to particular reactions.
- The molecular interactions taking place in cellular systems are inherently *discrete and stochastic processes*. This is a key feature of cellular systems that needs to be taken into account when describing their dynamics [9].
- It has been postulated that gene regulatory networks are organised in a *modular* manner in such a way that cellular processes arise from the orchestrated interactions between different genetic transcriptional units that can be considered separable modules [1].
- Spatial and geometric information must be represented in the system in order to describe processes involving *pattern formation*.

In this work we review recent advances in the use of the computational paradigm *membrane computing* or *P systems* as a formal methodology in *synthetic biology* for the specification and analysis on cellular system models according to the previously presented points.

K. Ambos-Spies, B. Löwe, and W. Merkle (Eds.): CiE 2009, LNCS 5635, pp. 430–436, 2009.

Stochastic P systems and modularity

P systems were introduced by G. Păun in 2000 as a computational abstraction of the structure and functioning of the *living cell* [11]. The main constituents of a generic P system are the following:

- A set of membranes representing compartments arranged in a hierarchical manner (membranes can contain other membranes) all of them embedded within a single membrane called *skin* that defines the system.
- Multisets of objects specifying molecular species distributed over the regions or compartments defined by membranes.
- Rewriting rules on multisets of objects associated specifically with each compartment and describing molecular interactions which determine the computation or evolution of the system.

For a detailed description we refer to [12]. Originally, the rewriting rules were applied according to a maximally parallel and nondeterministic strategy. More specifically, all the objects in every membrane that can evolve according to a rewriting rule must do so in a single step. If an object can evolve according to more than one rewriting rule the rule that is actually applied is chosen nondeterministically [11,12]. A considerable part of the research in this field has focused on the study of the computational universality and efficiency of the different proposed variants. Nevertheless, there is an emerging application of P systems as a framework for the specification and analysis of cellular systems [4,6,7,10,13,14,15,17,18,19,21,22].

In the following we will discuss *stochastic P systems*, a variant of the generic P system that satisfies the requirements presented previously to model cellular systems.

The original maximally parallel and nondeterministic semantics for the evolution of P systems was proved not suitable for reproducing the dynamics of cellular systems since it does not capture the different rates at which molecular interactions take place. In order to solve this problem stochastic semantics based on *Gillespie's theory of stochastic kinetics* [8] were introduced in the generic model to define *stochastic P system* [14]. This variant of P systems differs from the generic model in that a *stochastic kinetic constant* is associated specifically with each rewriting rule on multisets of objects. In particular, *boundary rules* [5] of the following form are used:

$$r: \; u\,[\,v\,]_i \xrightarrow{c} u'\,[\,v'\,]_i$$

where u, v, u', v' are multisets of objects that are replaced simultaneously outside and inside the corresponding compartment i represented with square brackets. The *kinetic stochastic constant* c associated with each rule is used to compute its propensity by multiplying it with the number of distinct possible combinations of objects in the left hand side of the rule. The sum of the propensities of all the rules is taken as the parameter of an exponential distribution which determines the time elapsed between rule applications. The rule that is applied is chosen

according to a multinomial distribution with parameters the normalised values of the rule propensities [8,14].

Stochastic P systems has been compared with other computational paradigms, like Petri nets and process algebra, that have also been adapted for their application to the modelling of cellular systems. Their similarities/differences and advantages/disadvantages were discussed [4,20].

As mention previously modularity is an interesting feature to study in cellular systems. Modularity can be explicitly represented in stochastic P systems by allowing the specification of the sets of rewriting rules associated with compartments as a combination of *instantiated P system modules* [21,22]. A P system module is a set of rewriting rules over multisets of variables representing objects, Var_O, whose stochastic kinetic constants are also variables Var_C. A module is identified with a name, Mod and is specified as $Mod(Var_O, Var_C)$. Then the instantiation of such module Mod with specific objects O_0, constants C_0 is specified as $Mod(O_0, C_0)$ and its rules are obtained from the rules in $Mod(Var_O, Var_C)$ by replacing all the occurrences of the variables Var_O and Var_C with the specific objects O_0 and stochastic kinetic constants C_0.

Note that the set of rewriting rules associated with a module Mod can be obtained by applying set union to simpler modules $Mod_1, \ldots, Mod_n$. In this way a nested hierarchical modular structure can be obtained.

Lattice Population P systems

Finally, a spatially distributed colony or tissue of cells can be represented using *lattice population P systems* introduced in [22]. Spatial and geometric information is introduced in P systems by using a finite regular geometrical *lattice*. The different types of cells in a colony or tissue are represented by individual stochastic P systems $\Pi_1, \ldots, \Pi_n$. Copies of these P systems are then distributed over the points of the geometrical lattice to represent the spatial distribution of cells in the colony or tissue. In this way each position $\mathbf{p}$ in the lattice has a specific P system associated with it that will be denoted $\Pi(\mathbf{p})$. The P systems distributed over the lattice can exchange objects using *translocation rules* of the following form which describe molecular processes like passive diffusion or active transport of molecular species.

$$r : [\, u \,]_l \overset{\mathbf{v}}{\Leftrightarrow} [\ ] \xrightarrow{c} [\ ]_l \overset{\mathbf{v}}{\Leftrightarrow} [\, u \,]$$

where u is a multiset of objects representing the molecular species that are to be diffused or transported, $\mathbf{v}$ is a vector and c is the corresponding kinetic stochastic constant. When rule r is applied in the P system located in position $\mathbf{p}$, $\Pi(\mathbf{p})$, the objects u are translocated from $\Pi(\mathbf{p})$ to the P system located in position $\mathbf{p} + \mathbf{v}$, $\Pi(\mathbf{p} + \mathbf{v})$.

In order to illustrate our approach we use the following example. Our genetic design consists of three colonies of bacteria arranged in a specific spatial distribution. These colonies of bacteria posses special genetic circuits that produce the temporal expression of a particular gene in a specific colony.

Table 1. The library of modules used in our example consisting in constitutive ($UnReg$), positive ($PosReg$) and negative ($NegReg$) gene regulation as well as protein degradation (Deg) and diffusion ($Diff$)

Module	**Rules**
$UnReg(\{G,P\},\{c_1\})$	$r_1 : [\ G\ ]_b \xrightarrow{c_1} [\ G+P\ ]_b$
$PosReg(\{A,G,P\},\{c_1,c_2,c_3\})$	$r_1 : [\ A+G\ ]_b \xrightarrow{c_1} [\ A.G\ ]_b$ $r_2 : [\ A.G\ ]_b \xrightarrow{c_2} [\ A+G\ ]_b$ $r_3 : [\ A.G\ ]_b \xrightarrow{c_3} [\ A.G+P\ ]_b$
$NegReg(\{R,G\},\{c_1,c_2\})$	$r_1 : [\ R+G\ ]_b \xrightarrow{c_1} [\ R.G\ ]_b$ $r_2 : [\ R.G\ ]_b \xrightarrow{c_2} [\ R+G\ ]_b$
$Deg(\{P\},\{c_1\})$	$r_1 : [\ P\ ]_b \xrightarrow{c_1} [\ ]_b$
$Diff(\{P\},\{c_1\})$	$r_1 : [\ P\ ]_b \overset{(1,0)}{\Leftrightarrow} [\ ] \xrightarrow{c_1} [\ ]_b \overset{(1,0)}{\Leftrightarrow} [\ P\ ]$ $r_2 : [\ P\ ]_b \overset{(-1,0)}{\Leftrightarrow} [\ ] \xrightarrow{c_1} [\ ]_b \overset{(-1,0)}{\Leftrightarrow} [\ P\ ]$ $r_3 : [\ P\ ]_b \overset{(0,1)}{\Leftrightarrow} [\ ] \xrightarrow{c_1} [\ ]_b \overset{(0,1)}{\Leftrightarrow} [\ P\ ]$ $r_4 : [\ P\ ]_b \overset{(0,-1)}{\Leftrightarrow} [\ ] \xrightarrow{c_1} [\ ]_b \overset{(0,-1)}{\Leftrightarrow} [\ P\ ]$

The three different types of bacteria forming the three colonies are specified by the P systems Π_1, Π_2 and Π_3, see Table 2. These P systems consist of a single membrane or compartment representing a bacterium. The corresponding set of rules associated with each P system is obtained as a combination of the P systems modules in Table 1. These modules represent the constitutive ($UnReg$), positive ($PosReg$) and negative ($NegReg$) regulation of genes as well as protein degradation (Deg) and diffusion ($Diff$). An additional P system Π_0 is used to represent a subvolume of the media between the different colonies.

The P systems from Table 2 are distributed over a rectangular lattice as shown in Figure 1. The leftmost colony consists of bacteria of type 1 represented by Π_1. These bacteria express constitutively $gene1$ whose protein product diffuses freely across the whole system. In the rightmost colony, described by Π_3, $gene2$ is regulated positively by $prot1$ producing the freely diffusible $prot2$. Finally, the colony in the centre of the lattice, specified by Π_2, possesses a genetic circuit according to which $prot1$ regulates positively $gene3$ whereas $prot2$ represses it.

Figure 2 shows the evolution over time of the molecules of $prot1, prot2$ and $prot3$ in the bacterium at position $(4,2)$ in the lattice presented in Figure 1. It can be observed that at around 50 minutes $prot1$ is present in the compartment producing the activation of $gene3$ and the accumulation of $prot3$. When $prot1$ reaches the rightmost colony, $prot2$ starts to be produced and it diffuses across the system. After a delay of around 300 minutes $prot2$ reaches our bacterium at position $(4,2)$ and represses the expression of $prot3$. Therefore our engineered design consisting in specific gene regulatory circuits and in a particular spatial

Table 2. The four different P systems, Π_0, Π_1, Π_2 and Π_3 representing a subvolume from the media, a bacterium from colony one, two and three respectively

P System	**Modules** [a]
Π_0	$Diff(\{prot1\}, \{0.01\})$ $Diff(\{prot2\}, \{0.01\})$
Π_1	$UnReg(\{gene1, prot1\}, \{10\})$ $Deg(\{prot1\}, \{0.0015\})$ $Deg(\{prot2\}, \{0.0015\})$ $Diff(\{prot1\}, \{0.01\})$ $Diff(\{prot2\}, \{0.01\})$
Π_2	$PosReg(\{prot1, gene3\}, \{1, 5, 5\})$ $NegReg(\{prot2, gene3\}, \{1, 0.001\})$ $Deg(\{prot1\}, \{0.0015\})$ $Deg(\{prot2\}, \{0.0015\})$ $Deg(\{prot3\}, \{0.03\})$ $Diff(\{prot1\}, \{0.01\})$ $Diff(\{prot2\}, \{0.01\})$
Π_3	$PosReg(\{prot1, gene2\}, \{1, 0.01, 10\})$ $Deg(\{prot1\}, \{0.0015\})$ $Deg(\{prot2\}, \{0.0015\})$ $Diff(\{prot1\}, \{0.01\})$ $Diff(\{prot2\}, \{0.01\})$

[a] The units of the constants associated with the modules are min^{-1}

Π_1 (0,4)	Π_1 (1,4)	Π_0 (2,4)	Π_0 (3,4)	Π_2 (4,4)	Π_2 (5,4)	Π_0 (6,4)	Π_0 (7,4)	Π_3 (8,4)	Π_3 (9,4)
Π_1 (0,3)	Π_1 (1,3)	Π_0 (2,3)	Π_0 (3,3)	Π_2 (4,3)	Π_2 (5,3)	Π_0 (6,3)	Π_0 (7,3)	Π_3 (8,3)	Π_3 (9,3)
Π_1 (0,2)	Π_1 (1,2)	Π_0 (2,2)	Π_0 (3,2)	Π_2 (4,2)	Π_2 (5,2)	Π_0 (6,2)	Π_0 (7,2)	Π_3 (8,2)	Π_3 (9,2)
Π_1 (0,1)	Π_1 (1,1)	Π_0 (2,1)	Π_0 (3,1)	Π_2 (4,1)	Π_2 (5,1)	Π_0 (6,1)	Π_0 (7,1)	Π_3 (8,1)	Π_3 (9,1)
Π_1 (0,0)	Π_1 (1,0)	Π_0 (2,0)	Π_0 (3,0)	Π_2 (4,0)	Π_2 (5,0)	Π_0 (6,0)	Π_0 (7,0)	Π_3 (8,0)	Π_3 (9,0)

Fig. 1. Spatial distribution of the three different bacterial colonies over a rectangular lattice

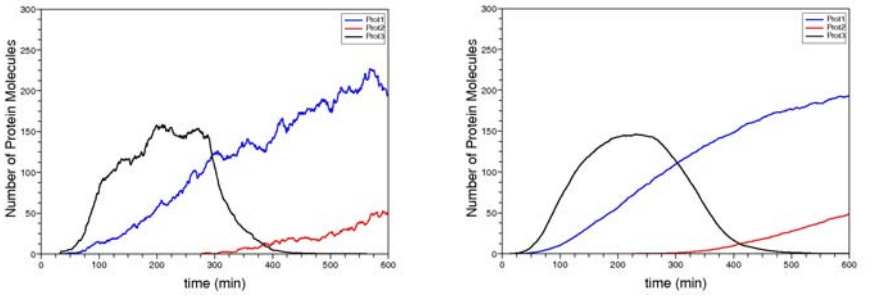

Fig. 2. Number of molecules of *prot*1, *prot*2 and *prot*3 in the bacterium at position (4,2) in single simulation (left) and the average over 100 simulations

distribution produces a transient expression of around 250 minutes of *prot*3 in the colony at the center of our system.

Acknowledgements

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

References

1. Alon, U.: Network motifs: theory and experimental approaches. Nature Reviews Genetics 8, 450–461 (2007)
2. Andrianantoandro, E., Basu, S., Karig, D., Weiss, R.: Synthetic biology: new engineering rules for an emerging discipline. Molecular Systems biology 2, 2006.0028 (2006)
3. Bernardini, F., Gheorghe, M., Romero-Campero, F.J., Walkinshaw, N.: A Hybrid Approach to Modeling Biological Systems. In: Eleftherakis, G., Kefalas, P., Păun, G., Rozenberg, G., Salomaa, A. (eds.) WMC 2007. LNCS, vol. 4860, pp. 138–159. Springer, Heidelberg (2007)
4. Bernardini, F., Gheorghe, M., Krasnogor, N.: Quorum Sensing P Systems. Theoretical Computer Science 371(1-2), 20–33 (2007)
5. Bernardini, F., Manca, V.: Dynamical aspects of P systems. BioSystems 70(2), 85–93 (2003)
6. 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(3), 377–385 (2008)
7. Bianco, L., Fontana, F., Manca, V.: P systems with reaction maps. International Journal of Foundations of Computer Science 17(1), 27–48 (2006)
8. Gillespie, D.T.: Stochastic Simulation of Chemical Kinetics. Annu. Rev. Phys. Chem. 58, 35–55 (2007)

9. Kaern, M., Elston, T.C., Blake, W.J., Collins, J.J.: Stochasticity in gene expression: from theories to phenotypes. Nature Reviews Genetics 6, 451–464 (2005)
10. 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)
11. Păun, G.: Computing with Membranes. Journal of Computer and System Sciences 61(1), 108–143 (2000)
12. Păun, G.: Membrane Computing: An Introduction. Springer, Heidelberg (2002)
13. Păun, A., Pérez-Jiménez, M.J., 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)
14. Pérez-Jiménez, M.J., 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)
15. Pescini, D., Besozzi, D., Mauri, G., Zandron, C.: Dynamical probabilistic P systems. International Journal of Foundations of Computer Science 17(1), 183–204 (2006)
16. Ptashne, M., Gann, A.: Genes and Signals. Cold Spring Harbor Laboratory Press (2002)
17. Romero-Campero, F.J., Pérez-Jiménez, M.J.: Modelling gene expression control using P systems: the Lac Operon, a case study. BioSystems 91(3), 438–457 (2008)
18. 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), 1–15 (2008)
19. 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)
20. Romero-Campero, F.J., Gheorghe, M., Ciobanu, G., Auld, J.M., Pérez-Jiménez, M.J.: Cellular modelling using P systems and process algebra. Progress in Natural Science 17(4), 375–383 (2007)
21. 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, July 12 - 16, pp. 331–338 (2008)
22. Romero-Campero, F.J., Twycross, J., Cámara, M., Bennett, M., Gheorghe, M., Krasnogor, N.: Modular Assembly of Cell Systems Biology Models Using P Systems. International Journal of Foundations of Computer Science (2008) (in press)