Modeling Signal Transduction Using P Systems

Andrei Păun[1], Mario J. Pérez-Jiménez[2], and Francisco J. Romero-Campero[2]

[1] Department of Computer Science/IfM, Louisiana Tech University
P.O. Box 10348, Ruston, LA 71272
apaun@latech.edu
[2] Department of Computer Science and Artificial Intelligence, University of Sevilla
Avda. Reina Mercedes s/n, 41012 Sevilla, Spain
{marper, fran}@us.es

Abstract. Cellular signalling pathways are fundamental to the control and regulation of cell behavior. Understanding of biosignalling network functions is crucial to the study of different diseases and to the design of effective therapies. In this paper we present P systems as a feasible computational modeling tool for cellular signalling pathways that takes into consideration the discrete character of the components of the system and the key role played by membranes in their functioning. We illustrate these cellular models simulating the epidermal growth factor receptor (EGFR) signalling cascade and the FAS–induced apoptosis using a deterministic strategy for the evolution of P systems.

1 Introduction

The complexity of biomolecular cell systems is currently the focus of intensive experimental research, nevertheless the enormous amount of data about the function, activity, and interactions of such systems makes necessary the development of models able to provide a better understanding of the dynamics and properties of the systems.

A model is an abstraction of the real-world onto a mathematical/computational domain that highlights some key features while ignoring others that are assumed to be not relevant. A good model should have four properties: relevance, computability, understandability, and extensibility, [22]. A model must be relevant capturing the essential properties of the phenomenon investigated, and computable so it can allow the simulation of its dynamic behavior, as well as the qualitative and quantitative reasoning about its properties. An understandable model will correspond well to the informal concepts and ideas of molecular biology. Finally, a good model should be extensible to higher levels of organizations, like tissues, organs, organisms, etc., in which molecular systems play a key role.

P systems are an unconventional model of computation inspired by the structure and functioning of living cells which takes into consideration the discrete character of the quantity of components of the system by using rewriting rules on multisets of objects, that represent chemical substances, and strings, that represent the organization of genes on the genome. The inherent randomness in

biological phenomena is captured by using stochastic strategies, [20]. We believe that P systems satisfy the above properties required for a good model.

Cellular signalling pathways are fundamental to the control and regulation of cell behavior. Understanding of biosignalling network functions is crucial to the study of different diseases and to the design of effective therapies. The characterization of properties about whole–cell functions requires mathematical/computational models that quantitatively describe the relationship between different cellular components.

Ordinary differential equations (ODEs) have been successfully used to model kinetics of conventional *macroscopic* chemical reactions. The approach followed by ODEs is referred as macroscopic chemistry since they model the average evolution of the concentration of chemical substances across the whole system. In this approach the change of chemical concentration over time is described for each chemical specie, implicitly assuming that the fluctuation around the average value of concentration is small relatively to the concentration. This assumption of homogeneity may be reasonable in some circumstances but not in many cases due to the internal structure and low numbers and non–uniform distributions of certain key molecules in the cell. While differential equations models may produce useful results under certain conditions, they provide a rather incomplete view of what is actually happening in the cell [2].

Due to the complexity of cellular signalling pathways, large number of linked ODEs are often necessary for a reaction kinetics model and the many interdependent differential equations can be very sensitive to their initial conditions and constants. Time delays and spatial effects (that play an important role in pathway behavior) are difficult to include in an ODE model [9], which are also very difficult to change and extend, because changes of network topology may require substantial changes in most of the basic equations [3].

Recently, different agent–based approaches are being used to model a wide variety of biological systems ([10], [12], [26]) and biological processes, including biochemical pathways [9].

The *microscopic* approach considers the molecular dynamics for each single molecule involved in the system taking into account their positions, momenta of atoms, etc. This approach is computationally intractable because of the number of atoms involved, the time scale and the uncertainty of initial conditions.

Our approach is referred as *mesoscopic* chemistry [25]. Like in the microscopic approach one considers individual molecules like proteins, DNA and mRNA, but ignores many molecules such as water and non-regulated parts of the cellular machinery. Besides, the position and momenta of the molecules are not modeled, instead one deals with the statistics of which reactions occur and how often. This approach is more tractable than microscopic chemistry but it provides a finer and better understanding than the macroscopic chemistry.

This paper is organized as follows. In the next section we present P systems as a framework for the specification of models of biosignalling cascades. A deterministic strategy for the evolution of P systems is described in Section 3. In Sections 4 and 5 a study of epidermal growth factor receptor (EGFR) signalling

cascade and of FAS–induced apoptotic signalling pathway are given. Finally, conclusions are presented in the last section.

2 P Systems: A Framework to Specify Biosignalling Cascades

In this paper we work with a variant of P systems of the form

$$\Pi = (O, L, \mu, M_1, M_2, \ldots, M_n, R_1, \ldots, R_n),$$

where:

- O is a finite alphabet of symbols representing objects (proteins and complexes of proteins);
- L is a finite alphabet of symbols representing labels for the compartments (membranes);
- μ is a membrane structure containing $n \geq 1$ membranes labeled with elements from L;
- $M_i = (w_i, l_i)$, $1 \leq i \leq n$, are pairs which represent the initial configuration of membrane i: $l_i \in L$ is its label, and $w_i \in O^*$ is the initial multiset.
- R_i, $1 \leq i \leq n$, are finite sets of rules associated with the membrane i which are of the form $u\,[\,v\,]_{l_i} \rightarrow u'\,[\,v'\,]_{l_i}$, where $u, v, u', v' \in O^*$ are finite multisets of objects and l_i is the label of membrane i.

Next, we discuss in more detail the rules that we will use in this paper, to model protein–protein interactions taking place in the compartmentalized structure of the living cell.

(a) *Transformation, complex formation and dissociation rules:*

$$\left.\begin{array}{l} [\ a\ ]_l \rightarrow [\ b\ ]_l \\ [\ a,\ b\ ]_l \rightarrow [\ c\ ]_l \\ [\ a\ ]_l \rightarrow [\ b,\ c\ ]_l \end{array}\right\} \quad \text{where } a, b, c \in O, \text{ and } l \in L$$

These rules are used to specify chemical reactions taking place inside a compartment of type $l \in L$; more specifically, they represent the transformation of a into b, the formation of a complex c from the interaction of a and b, and the dissociation of a complex a into b and c respectively.

(b) *Diffusing in and out:*

$$\left.\begin{array}{l} [\ a\ ]_l \rightarrow a\ [\ ]_l \\ a\ [\ ]_l \rightarrow [\ a\ ]_l \end{array}\right\} \quad \text{where } a \in O, \text{ and } l \in L$$

We use these types of rules when chemical substances move or diffuse freely from one compartment to another one.

(c) *Binding and debinding rules:*

$$\left.\begin{array}{l} a\,[\,b\,]_l \rightarrow [\,c\,]_l \\ \\ [\,a\,]_l \rightarrow b\,[\,c\,]_l \end{array}\right\} \quad \text{where } a, b, c \in O, \text{ and } l \in L$$

Using rules of the first type we can specify reactions consisting in the binding of a ligand swimming in one compartment to a receptor placed on the membrane surface of another compartment. The reverse reaction, debinding of substance from a receptor, can be described as well using the second rule.

(d) *Recruitment and releasing rules:*

$$\left.\begin{array}{l} a\,[\,b\,]_l \rightarrow c\,[\;\;]_l \\ \\ c\,[\;\;]_l \rightarrow a\,[\,b\,]_l \end{array}\right\} \quad \text{where } a, b, c \in O, \text{ and } l \in L$$

With these rules we represent the interaction between two chemicals in different compartments whereby one of them is recruited from its compartment by a chemical on the other compartment, and then the new complex remains in the latter compartment. In a releasing rule a complex, c, located in one compartment can dissociate into a and b, with remaining a in the same compartment as c, and b being released into the other compartment.

3 P Systems Using Deterministic Waiting Times Algorithm

In biological systems with a large number of molecules deterministic approaches are valid since the interactions between them follows the $\sqrt{n}$ law of physics, which states that randomness or fluctuation level in a system are inversely proportional to the square root of the number of particles.

Next, we present an *exact* deterministic strategy providing a semantic to the P systems defined before, that we will refer to as *deterministic waiting times algorithm*. It is based on the fact that *in vivo* chemical reactions take place in parallel in an asynchronous manner, i.e., different chemical reactions proceed at different reaction rates and the same reaction may also have different reaction rates at different times depending on the concentrations of reactants in the region.

In the deterministic waiting time strategy, the time necessary for a reaction to take place, called waiting time, is calculated and the rule or rules (chemical reaction) with the shortest waiting time is/are applied, changing the number of molecules in the respective compartments. In each step when there is a change in the number of a molecules in a compartment, the waiting time for the reactions "using" the changed molecule species has to be recalculated in that compartment.

By an exact deterministic method we mean that infinitesimal intervals of time are not approximated by Δt as it is the case in ODEs–based model, but we will

associate a waiting time, computed in a deterministic way, to each reaction and will use it to determine the order in which the reactions take place.

In our models biochemical reactions are used to describe the molecular interactions, and reversible complex formation reactions are frequent. In what follows we discuss how to compute mesoscopic rate constants from the macroscopic ones used in differential equations.

Our rules model reactions of the form:

$$A + B \underset{k_d}{\overset{k_a}{\rightleftharpoons}} C$$

This reversible reaction converges to an equilibrium in which the number of chemical species A, B and C remains constant. The equilibrium constant, K_{eq}, expresses the quantities of reactants A and B compared to complexes C once the equilibrium is reached; that is,

$$K_{eq} = \frac{[C]}{[A] \cdot [B]} \tag{1}$$

K_{eq} can also be computed using the association k_a and dissociation k_d rate constants:

$$K_{eq} = \frac{k_a}{k_d} \tag{2}$$

The association constant k_a determines the speed of the association reaction. It measures the number of chemicals A and B that form complexes C per mol and second. For the case of regulatory proteins the association rate constant k_a can be determined experimentally. K_{eq} can also be determined experimentally using (1) and therefore k_d can be computed using (2).

Alternatively, what it can be determined experimentally is Gibbs free energy ΔG, a notion from thermodynamics which measures the effort necessary for decomplexation. Gibbs free energy is related to the equilibrium constant K_{eq} as follows:

$$K_{eq} = \exp(\frac{-\Delta G}{R \cdot T}) \tag{3}$$

where $R = 1.9872\, cal\, mol^{-1}\, Kelvin^{-1}$ is the universal gas constant and T is the absolute temperature at which the experiments are performed.

Therefore from (2) and (3) the dissociation constant can be determined.

The rate constants k_a and k_d we have dealt with up to now are macroscopic, they do not depend on the actual number of molecules, but on concentration. Gillespie's algorithm and thus our approach uses mesoscopic rate constants referring to the actual number of molecules and they are determined from their macroscopic counterparts as follows:

$$c_a = \frac{k_a}{A \cdot V}, \qquad c_d = k_d$$

where $A = 6.023 \cdot 10^{23}$ is Avogadro's number and V is the cell volume. Note that we assume the cell volume to be constant while ignoring cell growth.

Given a P system, in this strategy each rule r (representing a chemical reaction) in each membrane m has associated a *velocity*, v_r, obtained by multiplying the mesoscopic rate constant c_r by the multiplicities of the reactants according to the mass action law. Then we compute the waiting time for the first execution of the rule r as $\tau_r = \frac{1}{v_r}$ and return the triple (τ_r, r, m).

Next, we give a detailed description of the *deterministic waiting times algorithm* providing the semantic of our P systems–based model:

- **Initialization**
  - ⋆ set time of the simulation $t = 0$;
  - ⋆ for every rule r associated with a membrane m in μ compute the triple (τ_r, r, m) by using the procedure described before; construct a list containing all such triples;
  - ⋆ sort the list of triple (τ_r, r, m) according to τ_r (in an ascendent order);
- **Iteration**
  - ⋆ extract the first triple, (τ_r, r, m) from the list (if there are several rules with the minimum waiting times, then we select all these rules);
  - ⋆ set time of the simulation $t = t + \tau_r$;
  - ⋆ update the waiting time for the rest of the triples in the list by subtracting τ_r;
  - ⋆ apply the rule(s) r only once updating the multiplicities of objects in the membranes affected by the application of the rule;
  - ⋆ for each membrane m' affected by the application of the rule(s) r, recalculate the waiting times of the rules which are in m';
  - ⋆ for each such rule, compare the new waiting times with the existing ones, and keep the smallest one among the two;
  - ⋆ sort the list of the new triples according to the waiting time;
  - ⋆ iterate the process.
- **Termination**
  - ⋆ Repeat the process until the time of the simulation t reaches or exceeds a preset maximal time of simulation.

Note that in this algorithm every rule in each membrane has a waiting time computed in a deterministic way that is used to determine the order in which the rules are executed. It is also worth mentioning that in this method the time step varies across the evolution of the system and it is computed in each step depending on the current state of the system.

This strategy has been implemented using Scilab, a scientific software package for numerical computations providing a powerful open computing environment for engineering and scientific applications [31]. This tool is available from [32].

4 Modeling EGFR Signalling

The epidermal growth factor receptor (EGFR) is provably the best understood receptor system, and computational models have played an important role in its elucidation. It seems clear that cells process the information before passing it to

the nucleus. The many different control points in the EGFR signalling pathway make it an excellent system for investigating how cells process contextual information. Computational models can play a crucial role for understanding this process [29].

In this section we study the EGFR signalling cascade where the deterministic waiting times algorithm is suitable for describing its evolution.

The epidermal growth factor receptor (EGFR) was the first found to have tyrosine–kinase activity and has been used in pioneering studies of biological processes such as receptor–mediated endocytosis, oncogenesis, mitogen–activated–protein–kinase (MAPK) signalling pathways, etc. [29].

Binding of the epidermal growth factor (EGF) to the extracellular domain of EGFR induces receptor dimerization and autophosphorylation of intracellular domains. Then a multitude of proteins are recruited starting a complex signalling cascade and the receptor follows a process of internalization, ubiquitination, and degradation in endosomals.

In our model we consider two marginal pathways and two principal pathways starting from the phosphorylated receptor.

In the first marginal pathway phospholipase C-γ (PLC_γ) binds to the phospholyrated receptor, then it is phosphorylated (PLC^*_γ) and released into the cytoplasm where it can be translocated to the cell membrane or desphosphorylated. In the second marginal pathway the protein PI3K binds to the phospholyrated receptor, then it is phosphorylated ($PI3K^*$) and released into the cytoplasm where it regulates several proteins that we do not include in our model.

Both principal pathways lead to activation of Ras-GTP. The first pathway does not depend on the concentration of the Src homology and collagen domain protein (Shc). This pathway consists of a cycle where the proteins growth factor receptor-binding protein 2 (Grb2) and Son of Sevenless homolog protein (SOS) bind to the phosphorylated receptor. Later the complex Grb2-SOS is released in the cytoplasm where it dissociates into Grb2 and SOS.

In the other main pathway Shc plays a key role, it binds to the receptor and it is phosphorylated. Then either Shc^* is released in the cytoplasm or the proteins Grb2 and SOS binds to the receptor yielding a four protein complex (EGFR-EGF2*-Shc*-Grb2-SOS). Subsequently, this complex dissociates into the complexes Shc^*-Grb2-SOS, Shc^*-Grb2 and Grb2-SOS which in turn can also dissociate to produce the proteins Shc^*, Grb2 and SOS.

Finally, Ras-GTP is activated by these two pathways and in turn it stimulates the Mitogen Activated Protein (MAP) kinase cascade by phosphorylating the proteins Raf, MEK and ERK. Subsequently, phosphorylated ERK regulates several cellular proteins and nuclear transcription factors that we do not include in our model.

There exist *cross-talks* between different parts and cycles of the signalling cascade which suggests a strong robustness of the system.

In Figure 1 it is shown a detailed graphical representation of the signalling pathway that we model in this paper.

Next, we present a P system–based model of the biosignalling cascade described above.

Our model consists of more that 60 proteins and complexes of proteins and 160 chemical reactions. We will not give all the details of the model. A complete description of $\mathbf{\Pi}_{EGF}$ with some supplementary information is available from the web page [32]. In what follows we give an outline of our model.

Let us consider the P system

$$\mathbf{\Pi}_{EGF} = (O, \{e, s, c\}, \mu, (w_1, e), (w_2, s), (w_3, c), \mathcal{R}_e, \mathcal{R}_s, \mathcal{R}_c),$$

where:

• **Alphabet:** In the alphabet O we represent all the proteins and complexes of proteins that take part in the signalling cascade simulated. Some of the objects from the alphabet and the chemical compounds that they represent are listed below.

Object	**Protein or Complex**
EGF	Epidermal Growth Factor
EGFR	Epidermal Growth Factor Receptor
EGFR-EGF_2	Dimerisated Receptor
EGFR-EGF_2^*-Shc	EGFR-EGF_2^* and Shc complex
⋮	⋮
MEK	Mitogenic External Regulated Kinase
ERK	External Regulated Kinase

• **Membrane Structure:** In the EGFR signalling cascade there are three relevant regions, namely the *environment*, the *cell surface* and the *cytoplasm*. We represent them in the membrane structure as the membranes labeled with e for the environment, s for the cell surface, and c for the cytoplasm. The skin of the structure is the environment, the cell surface is the son of the environment and the father of the cytoplasm.

• **Initial Multisets:** In the initial multisets we represent the initial number of molecules of the chemical substances in the environment, the cell surface, and the cytoplasm. These estimations has been obtained from [13,24].

$$\begin{aligned} w_e &= \{EGF^{20000}\} \\ w_s &= \{EGFR^{25000}, Ras\text{-}GDP^{20000}\} \\ w_c &= \{Shc^{25000}, PLC_\gamma^{15000}, PI3K^{5000}, SOS^{4000}, Grb2^{8000}, TP_1^{10000}, TP_2^{45000}, \\ &\quad TP_3^{45000}, TP_4^{12500}, Raf^{8000}, MEK^{40000}, ERK^{40000}, P_1^{8000}, P_2^{8000}, P_3^{30000}\} \end{aligned}$$

• **Rules:** Using rules we model the 160 chemical reactions which form the signalling cascade.

As it can be seen in the initial multisets specified before, in the system of the EGFR signalling cascade the number of molecules is quite large, hence as a consequence of the $\sqrt{n}$ law important fluctuations and stochastic behavior are

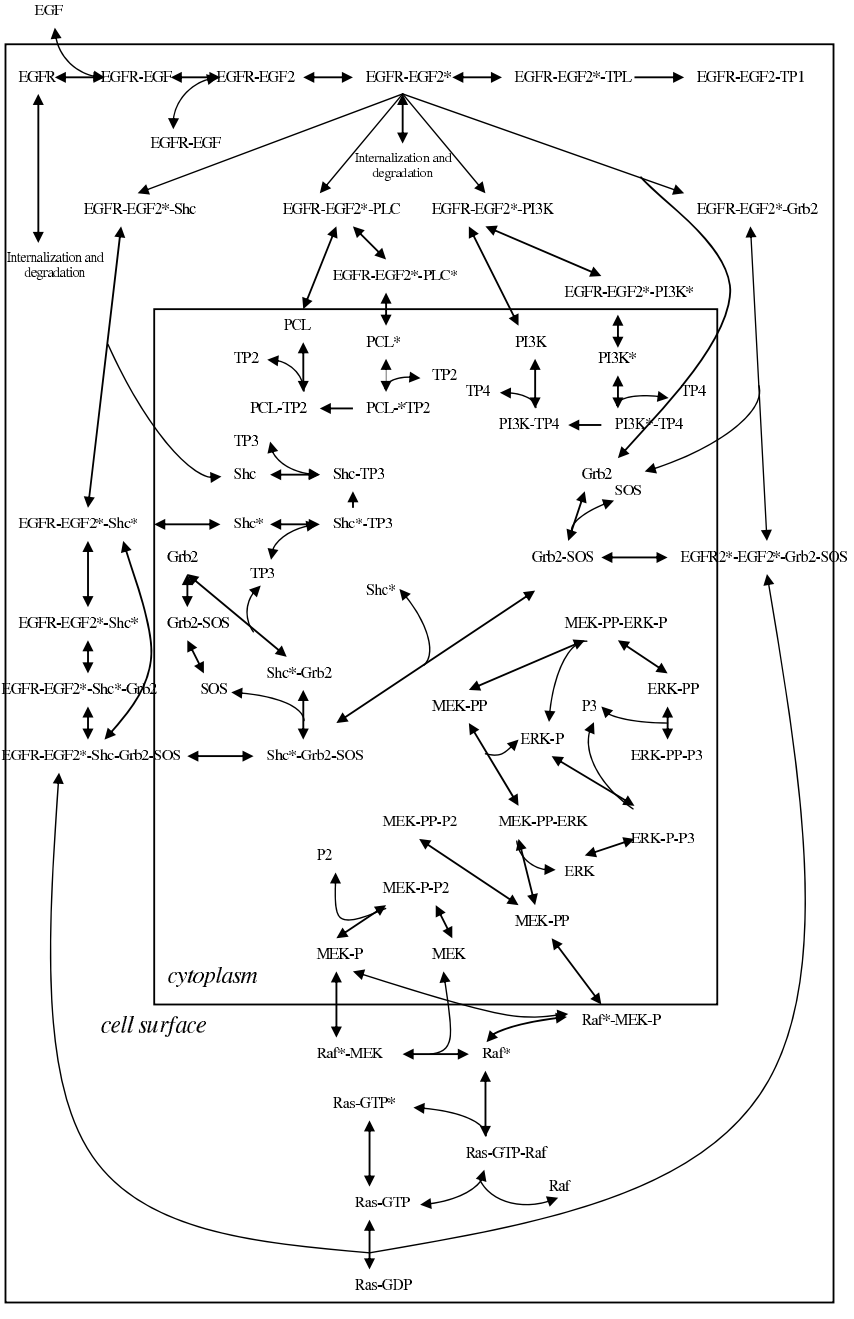

Fig. 1. EGFR Signalling Cascade

not expected in the evolution of the system. Because of this we have chosen the deterministic waiting times algorithm as the strategy for the evolution of the P system $\mathbf{\Pi}_{EGF}$.

Next, we show two examples of rules of the system.

The set of rules associated with the environment, $\mathcal{R}_e$, consists only of one rule r which models the binding of the signal, EGF, to the receptor $EGFR$.

$$EGF\,[\,EGFR\,]_s \;\rightarrow\; [\,EGF\text{-}EGFR\,]_s\;, c_r$$

The meaning of the previous rule is the following: the object EGF in the membrane containing the membrane with label s (the environment), and the object $EGFR$ inside the membrane with label s (the cell surface) are replaced with the object $EGFR$-EGF in the membrane with label s; this object represents the complex receptor-signal on the cell surface. We associate the mesoscopic rate constant c_r, which measures the affinity between the signal and the receptor.

The deterministic waiting times algorithm is used in the evolution of the system and the waiting time associated to this rule will be computed using the next formula:

$$\tau_r = \frac{1}{c_r \cdot |EGF| \cdot |EGFR|}$$

One example from the set of rules $\mathcal{R}_s$ associated with the cell surface is the rule r' concerning to the dimerisation of the receptor, that is the formation of a complex consisting of two receptors:

$$[\,EGFR,\, EGFR\,]_s \;\rightarrow\; [\,EGFR_2\,]_s\;, c_{r'}$$

When this rule r' is executed two objects $EGFR$ representing receptors are replaced with one object $EGFR_2$, representing a complex formed with two receptors, in the membrane with label s, the cell surface. The mesoscopic rate constant $c_{r'}$ is used to computed the waiting time:

$$\tau_{r'} = \frac{1}{c_{r'} \cdot |EGFR|^2}$$

4.1 Results and Discussions

Using Scilab we ran some experiments; in what follows we present some of the results obtained.

In Figure 2 it is depicted the evolution of the number of autophosphorylated receptors and in Figure 3 the number of doubly phosphorylated MEK (Mitogen External Kinase), one of the target proteins of the signalling cascade that regulates some nuclear transcription factors involved in the cell division.

Note that the activation of the receptor is very fast reaching its maximum within the first 5 seconds and then it decays fast to very low levels; on the other hand the number of doubly phosphorylated MEK is more sustained around 3 nM. These results agree well with empirical observations, see [13,24].

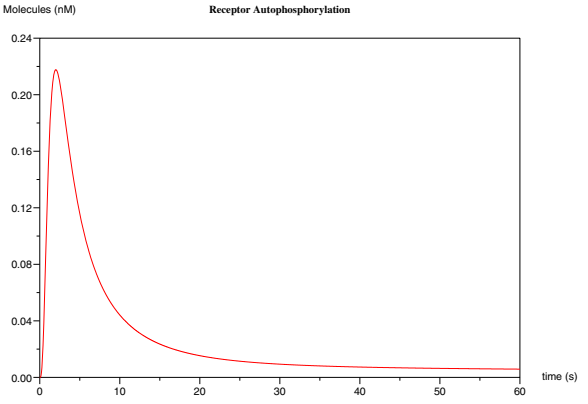

Fig. 2. Autophosphorylated EGFR evolution

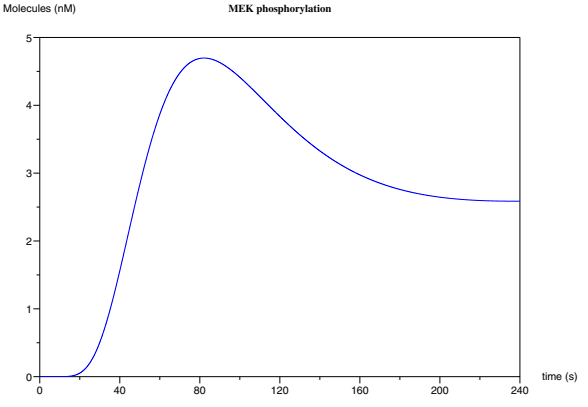

Fig. 3. Doubly phosphorylated MEK evolution

In tumors it has been reported an over expression of EGF signals in the environment and of EGFR receptors on the cell surface of cancerous cells. Here we investigate the effect of different EGF concentrations and number of receptors on the signalling cascade.

First, we study the effect on the evolution of the number of autophosphorylated receptors and doubly phosphorylated MEK of a range of signals, EGF, from 100 nM to 2000 nM.

In Figure 4, it can be seen that the receptor autophosphorylation is clearly concentration dependent showing different peaks for different number of signals in the environment. According to the variance in the receptor activation it is intuitive to expect different cell responses to different EGF concentrations. Here we will see that this is not the case.

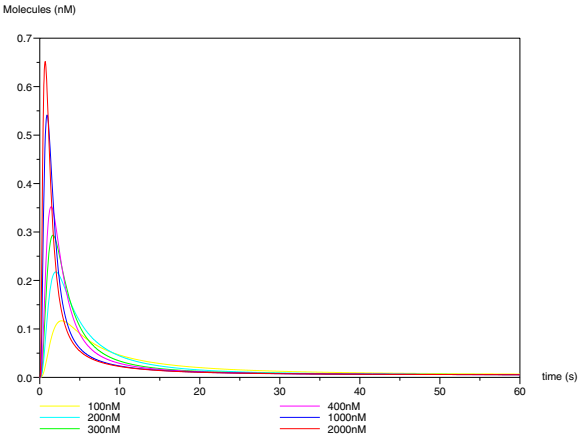

Fig. 4. Receptor autophosphorylation for different environmental EGF concentrations

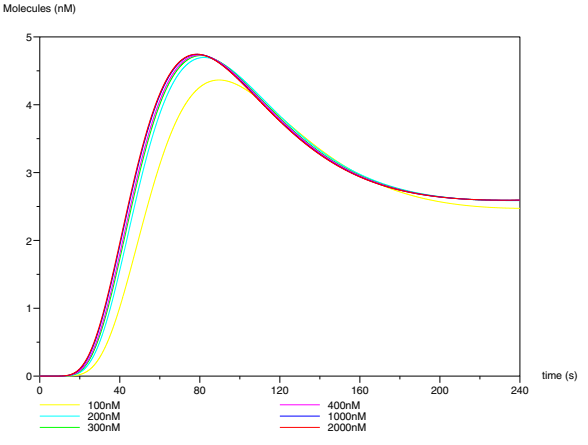

Fig. 5. MEK phosphorylation for different environmental EGF concentrations

From Figure 5 we can observe that the number of doubly phosphorylated MEK does not depend on the number of signals in the environment. That is, the perturbation of EGF level in the environment has no impact on MEK activation. So, the system is not sensitive to increases in EGF level in the environment (*Sensitivity analysis* is a mathematical technique term associated with the use of a computational model to predict the effects of the variation of a single component in the model).

This shows the surprising robustness of the signalling cascade with regard to the number of signals from outside due to EGF concentration. The signal is either attenuated or amplified to get the same concentration of one of the most relevant kinases in the signalling cascade, MEK. Note that after 100 seconds, when the

response gets sustained, the lines representing the response to different external EGF concentrations are identical.

Now we analyze the effect on the dynamics of the signalling cascade of different numbers of receptors on the cell surface.

In Figure 6, it is shown the evolution of the number of doubly phosphorylated MEK when there is 100 nM and 1000 nM of receptors on the cell surface. Note that now the response is considerably different; the number of activated MEK is greater when there is an over expression of receptors on the cell surface. As a consequence of this high number of activated MEK the cells will undergo an uncontrolled process of proliferation. Thus, Figure 6 shows the sensitivity of MEK activation to increases in EGFR level on the cell surface.

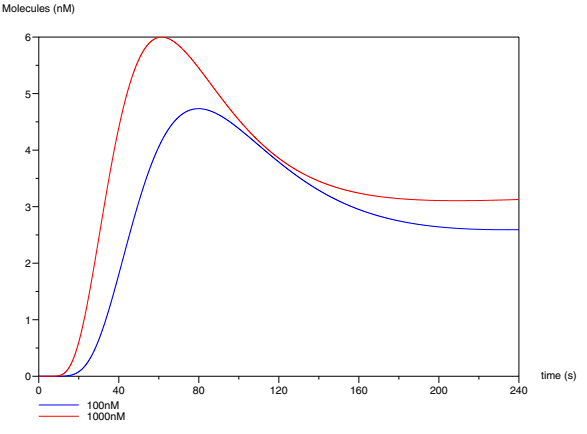

Fig. 6. MEK phosphorylation for different number of receptors

The key role played by the over expression of EGFR on the uncontrolled growth of tumors has been reported before, and, as a consequence of this, EGFR is one of the main biological targets for the development of novel and successful therapies against cancer and continue to be a source of discoveries about the cell signalling mechanisms involved in development, tissue homeostasis, and disease [28,14].

There are different strategies to inhibit the over expression of EGFR on the cell surface but the most developed ones are the *monoclonal antibodies* (that bind the external domain of the receptor competing against their natural ligands), and the *molecules with low molecular weight* (that inhibit the tyrosine–kinase activity of the receptor at the intracellular level).

Finally, we stress that for this system we have used a deterministic approach obtaining results that agree well with experimental data. This is not always the case, for instance in [20] a system is shown (the Quorum Sensing system in Vibrio Fischeri) where a stochastic approach is necessary to describe properly its behavior.

5 Modelling FAS–Apoptosis

There are basically two mechanisms of cell death, *necrosis* and *apoptosis*. Necrosis is a form of cell death that usually occurs when cells are damaged by injury. A disruption of the cell membrane is produced and intracellular materials are released. In contrast to necrosis, apoptosis is carried out in an ordered sequence of events that culminates in the suicide of the cell, and without releasing intracellular materials from the dying cells.

The term *apoptosis* (also known as *programmed cell death*) was coined by Kerr, Wyllie and Currie [8] as a means of distinguishing a morphologically distinctive form of cell death which was associated with normal physiology.

Apoptosis occurs during organ development, it plays an important role in cellular homeostasis [11], and it is a cellular response to a *cellular insult* that starts a cascade of apoptotic signals, both intracellular and extracellular, which converge on the activation of a group of apoptotic–specific proteases called caspases. The apoptotic mechanism include condensation of cell contents, DNA fragmentation into nucleosomal fragments, nuclear membrane breakdown, and the formation of apoptotic bodies that are small membrane–bound vesicles phagocytosed by neighboring cells [15]. Apoptosis protects the rest of the organism from a potentially harmful agent and disregulation of apoptosis can contribute to the development of autoimmune diseases and cancers. Apoptosis can also be induced by anticancer drugs, group factor deprivation, and irradiation.

The family of proteases that mediates apoptosis is divided into two subgroups. The first group consists of caspase 8, caspase 9, and caspase 10, and they function as initiators of the cell death process. The second group contains caspase 3, caspase 6, and caspase 7, and they work as effectors. The other effector molecule in apoptosis is Apaf-1, which, together with cytochrome c, stimulates the processing of pro-caspase 9 to the mature enzyme.

The other regulators of apoptosis are the Bcl2 family members, divided into three subgroups based on their structure. Members of the first subgroup, represented by Bcl2 and Bcl-xL, have an anti-apoptotic function. The second subgroup, represented by Bax and Bak, and the third subgroup, represented by Bid and Bad, are pro-apoptotic molecules.

Apoptotic death can be triggered by a wide variety of stimuli. Among the more studied death stimuli are DNA damage which in many cells leads to apoptotic death via a pathway dependent on p53, and the signalling pathways for FAS-induced apoptosis that was shown to be one of the most relevant processes for understanding and combating many forms of human diseases such as cancer, neurodegenerative diseases (Parkinson's disease, Alzheimer, etc.), AIDS and ischemic stroke.

Fas (also called CD95 or APO–1) is a cell surface receptor protein with an extracellular region, one transmembrane domain, and an intracellular region. Fas belongs to the tumor necrosis factor/nerve growth factor (TNT/NGF) cytokine receptor family. Activation of Fas through binding to its ligands, induces apoptosis in the Fas bearing cell. Fas induced–apoptosis starts from the Fas ligand

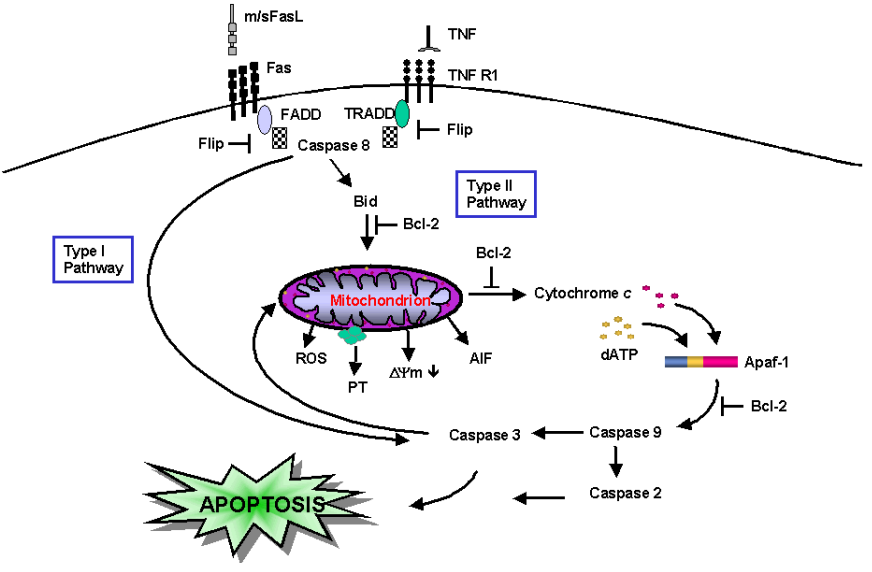

Fig. 7. FAS signalling pathways, from [1]

binding to Fas receptors and ends in the fragmentation of genomic DNA, which is used as a hallmark of apoptosis.

Fas ligands usually exist as trimers and bind and activate their receptors by inducing receptor trimerisation. This creates a clustering of Fas that is necessary for signalling. In its intracellular region, Fas contains a conserved sequence called a *death domain*. Activated receptors recruit adaptor molecules (such as FADD, Fas–associating protein with death domain) which interacts with the death domain on the Fas receptor and recruit procaspase 8 to the receptor complex, where it undergoes autocatalytic activation cleaving and releasing active caspase 8 molecules intracellularly. Activated caspase 8 can activate caspase 3 through two different pathways that have been identified by Scaffidi et al. [23], and are referred to as type I (*death receptor pathway*) and type II (*mitochondrial pathway*), where caspases play a crucial role for both the initiation and execution apoptosis.

The pathways diverge after activation of initiator caspases and converge at the end by activating executor caspases. In the type I pathway, initiator caspase (caspase 8) cleaves procaspase 3 directly and activates executor caspase (caspase 3).

In the type II pathway, a more complicated cascade is activated involving the disruption of mitochondrial membrane potential and it is mediated by Bcl2 family proteins that regulate the passage of small molecules which activate caspase cascades through the mitochondrial transition pore. More specifically (see Figure 8), caspase 8 cleaves Bid (Bcl2 interacting protein) and its COOH–terminal part translocates to mitochondria where it triggers cytochrome c release. The released cytochrome c bind to Apaf–1 (apoplectic protease activating factor)

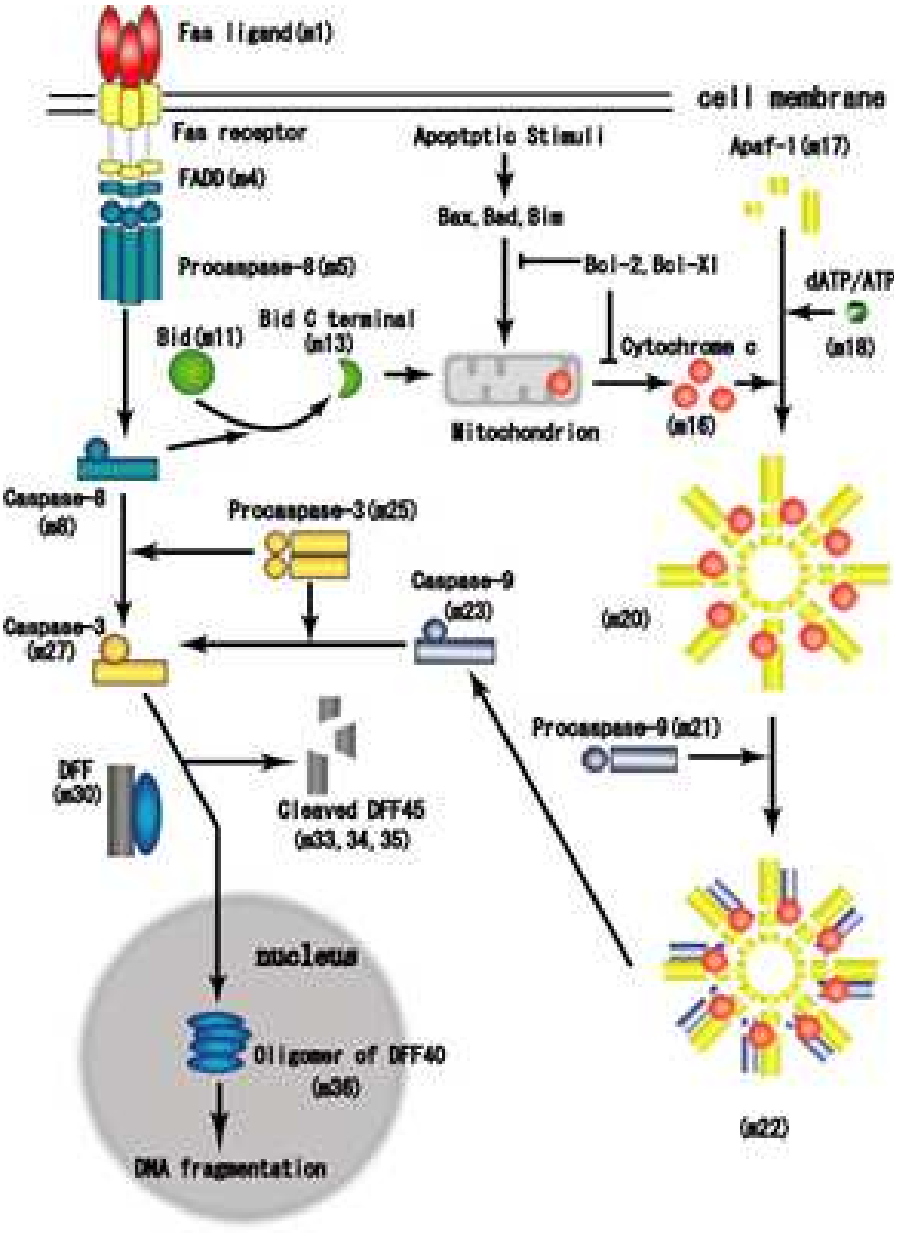

Fig. 8. Details of FAS signalling pathways, from [15]

togegther with dATP and procaspase 9 and activate caspase 9. The caspase 9 cleaves procaspase 3 and activates caspase 3.

The executor caspase 3 cleaves DFF (DNA fragmentation factor) in a heterodimeric factor of DFF40 and DFF45. Cleaved DFF45 dissociates from DFF40,

inducing oligomerisation of DFF40. The active DFF40 oligomer causes the internucleosomal DNA fragmentation.

Despite many molecular components of these apoptotic pathways have been identified, a better understanding of how they work together into a consistent network is necessary. A way to understand complex biological processes, in general, and the complex signalling behaviour of these pathways, in particular, is by modelling them in a computational framework and simulating them in electronic computers.

In [6] the two pathways activated by FAS starting with the stimulation of FASL (FAS ligand) until the activation of the effector caspase 3, have been modelled using ordinary differential equations in which biochemical reactions where used to describe molecular interactions.

In this section we present a P system using a deterministic waiting times algorithm for modelling FAS induced apoptosis, implementing all the rules described in [6] for both pathways.

Our model consists of 53 proteins and complexes of proteins and 99 chemical reactions. We will not give all the details of the model. A complete description of $\mathbf{\Pi}_{FAS}$ with some supplementary information can be found in the web page [32]. In what follows we give an outline of our model.

Let us consider the P system

$$\mathbf{\Pi}_{FAS} = (O, \{e, s, c, m\}, \mu, (w_1, e), (w_2, s), (w_3, c), (w_4, m), \mathcal{R}_e, \mathcal{R}_s, \mathcal{R}_c, \mathcal{R}_m)$$

where:

• **Alphabet:** In the alphabet O we represent all the proteins and complexes of proteins that take part in the signalling cascade simulated. Some of the objects from the alphabet and the chemical compounds that they represent are listed below.

Object	**Protein or Complex**
FAS	Fas protein
FASL	Fas Ligand
FADD	Fas–associating protein with death domain
⋮	⋮
Apaf	Apoptotic protease activating factor
Smac	Second mitochondria–derived activator of caspase
XIAP	X–linked inhibitor of apoptosis protein

• **Membrane Structure:** In the FAS signalling pathways there are four relevant regions, namely the *environment*, the *cell surface*, the *cytoplasm* and the *mitochondria*. We represent them in the membrane structure as the membranes labeled with: e for the environment, s for the cell surface, c for the cytoplasm, and m for the *mitochondria*. The skin of the structure is the environment, the cell surface is the son of the environment, the father of the cytoplasm, and the grandfather of the mitochondria.

• **Initial Multisets:** In the initial multisets we represent the initial number of molecules of the chemical substances in the environment, the cell surface, the cytoplasm, the mitochondria. These estimations has been obtained from [6].

$$\begin{aligned} w_e &= \{FASL^{12500}\} \\ w_s &= \{FAS^{6023}\} \\ w_c &= \{FADD^{10040}, CASP8^{20074}, FLIP^{48786}, CASP3^{120460}, Bid^{15057}, \\ &\quad Bax^{50189}, XIAP^{18069}, Apaf^{60230}, CASP9^{12046}\} \\ w_m &= \{Smac^{60230}, Cyto.c^{60230}, Bcl2^{45172}\} \end{aligned}$$

• **Rules:** Through the rules we model the 99 chemical reactions which form the signalling pathways. The rules can be found in [5] and they are described in our model as in the case of the system $\mathbf{\Pi}_{EGFR}$ (with different rules in the alternative cases of type II pathway in next subsection).

The set of rules associated with the environment, $\mathcal{R}_e$, consists only of one rule r_1 which models the binding of the FAS ligand to the receptor FAS.

$$FASL\ [\ FAS\ ]_s \rightarrow [\ FASC\ ]_s\ , c_{r_1}$$

The meaning of the previous rule is the following: the object $FASL$ in the membrane containing the membrane with label s (the environment), and the object FAS inside the membrane with label s (the cell surface) are replaced with the object $FASC$ in the membrane with label s; this object represents the complex receptor-signal on the cell surface. We associate the kinetic constant k_1, which measures the affinity between the signal and the receptor.

The deterministic waiting times algorithm is used in the evolution of the system and the waiting time associated to this rule will be computed using the next formula:

$$\tau_{r_1} = \frac{1}{c_{r_1} \cdot |EGF| \cdot |EGFR|}$$

5.1 Results and Discussions

We implemented in Java a preliminary *simulator* for the P system. It accepts as input an SBML (Systems Biology Markup Language) file containing the rules to be simulated and initial concentrations for the molecules in the system.

We compared our results with both the experimental data and with the ODEs simulation data reported in [6].

One of the major proteins in the pathway, caspase 3 was compared to the experimental data. In the ODEs simulation, caspase 3 was activated at 4 hours, and it was considered close to the experimental results where it was obtained that it activated at 6 hours (see Figure 9).

The same pathway is modeled in the membrane computing framework using the same reactions and initial conditions. The caspase 3 activation dynamics is studied when Bcl2 is at baseline value. Caspase 3 is activated in our simulator

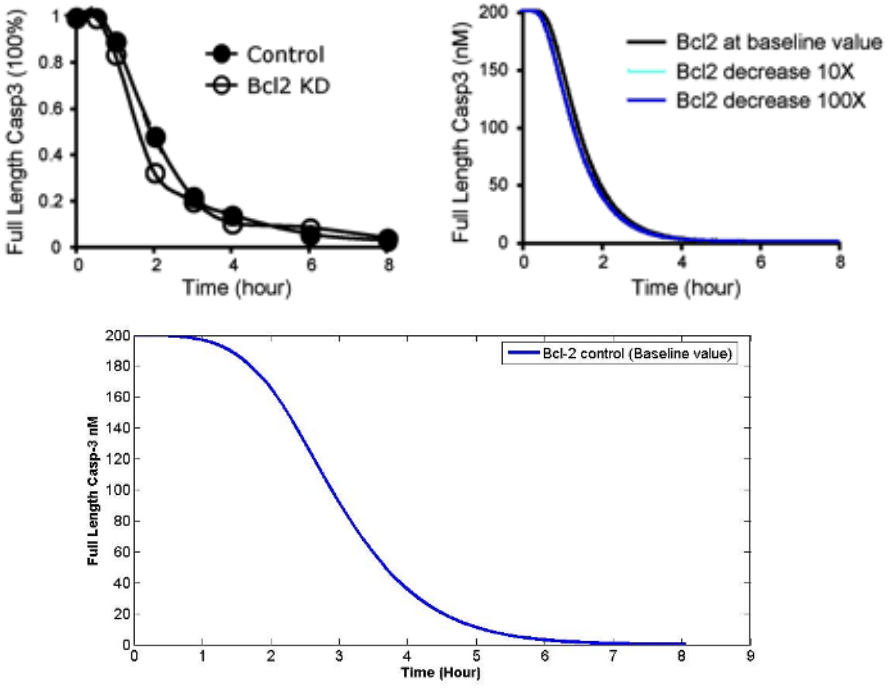

Fig. 9. Comparison between experimental data (top left, from [6]), previous ODE simulation data (top right, [6]) and the P system simulation data (down)

after about 7 hours which is a very good approximation of the experimental data and it improves the results obtained in the ODEs simulation [6].

There are cells (as thymocytes and fibroblasts) which are not sensitive to Bcl2 over expression as described in [23]. In these cells caspase 8 directly activates caspase 3.

Scaffidi et al. has suggested in [23] that the type of pathway activated by Fas is chosen based on the concentration of caspase 8 generated in active form following FASL binding. If the concentration of activated caspase 8 is high, then the caspase 3 is activated directly, on the other hand, if the concentration of activated caspase 8 is low, the type II pathway is chosen so that the system is amplifying the death signal through the mitochondria to be able to induce the cell death.

To check this hypothesis, the active caspase 8 formation is increased by having the initial concentration of caspase 8 set to a value 20 times greater than its baseline value while everything else was kept the same in the system. We performed the same simulation with the increase in caspase 8 initial concentration, and this resulted in faster caspase 3 activation also in our simulation; this agrees well with the results obtained in [23].

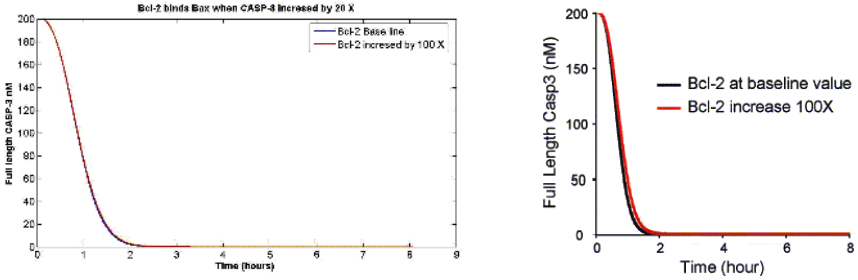

Fig. 10. Left – the P system simulation, right – the ODE simulation, from [6], for the change in caspase 8 initial concentration

The Bcl2 concentration is also increased 100 times to test the sensitivity of caspase 3 activation to Bcl2. Figure 10 shows that the caspase 3 activation is not sensitive to increases in Bcl2 concentration, when pathway of type I is chosen.

Bcl2 is known to block the mitochondrial pathway; however, it is not clear the mechanism through which Bcl2 can block the pathway of type II. Next, we analyze the caspase 3 activation kinetics in this type of pathway by considering different mechanisms to block the mitochondrial pathway suggested in [4], [16] and [27]: Bcl2 might bind with (a) Bax, (b) Bid, (c) tBid, or (d) bind to both Bax and tBid.

We design four different P systems having the rules:

- $r_1, \ldots, r_{95}, r_{96}, r_{97}$ for modeling the case (a).
- $r_1, \ldots, r_{95}, r'_{96}, r'_{97}$ for modeling the case (b).
- $r_1, \ldots, r_{95}, r''_{96}, r''_{97}$ for modeling the case (c).
- $r_1, \ldots, r_{97}, r_{98}, r_{99}$ for modeling the case (d).

All the other rules remain the same for all the cases (see [5] for details).

Let us note that this example shows the modularity of P systems–based model: small behavioral changes in the biosignalling cascade causes small changes in the designs of the P systems.

The dynamics of caspase 3 activation is studied by varying the Bcl2 concentration 10 times or 100 times the baseline value. It was concluded that Bcl2 binding to both Bax and tBid is the most efficient mechanism for the pathway in comparison with the results obtained for the cases (a), (b) or (c). The same conclusions were obtained also after using our simulator for all the previous changes in the pathway.

Figure 11 shows only the case (d) as a comparison between the ODE simulator and the P system simulator. It can seen the sensitivity of the caspase 3 activation to increases in Bcl2 level, when Bcl2 is able to bind to both Bax and tBid, and when mitochondrial pathway is selected.

Next table presents a summary about the sensitivity analysis of caspase 3 activation to over expression of Bcl2 in function of the pathway selected.

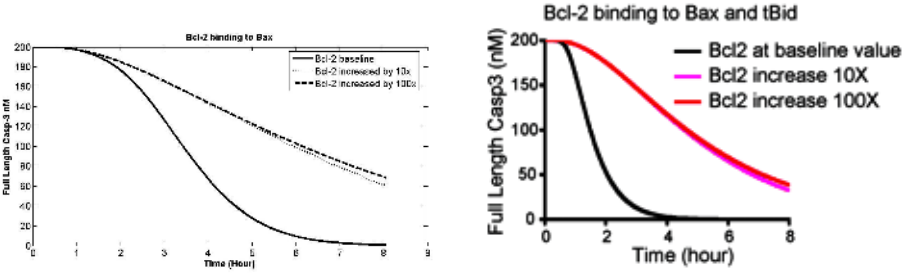

Fig. 11. Left – the P system simulation, right – the ODE simulation, from [6]

	Activation Caspase 3 (with over expression of Bcl2)
Type I (death receptor pathway)	*Insensitivity*
Type II (mitochondrial pathway)	*Sensitivity*

6 Conclusions

In this paper we have presented P systems as a new computational modeling tool to study the dynamic behavior of integrated signalling systems through a mesoscopic chemistry approach.

P systems are also a general specification of the biological phenomena that can be evolved using different strategies/algorithms. A *deterministic waiting times algorithm* has been introduced, and it is based on the fact that *in vivo* chemical reactions take place in parallel and in an asynchronous manner.

That strategy has been illustrated with the simulation of two relevant biological phenomena: the EGFR signalling cascade and the signalling pathways for FAS–induced apoptosis. In the line of [29] we think that the success of using P systems–based model for simulating biosignalling cascades can be a guide to combining models and experiments to understand complex biological processes as integrated systems.

Our results show a good correlation with the experimental data reported in the literature and with simulators based on ODEs. So, they support the reliability of P systems as computational modeling tools to produce postdiction, and perhaps they will be able to produce plausible predictions.

Acknowledgement

The author wishes to acknowledge the support of the project TIN2005-09345-C04-01 of the Ministerio de Educación y Ciencia of Spain, co–financed by FEDER funds, and of the project of Excellence TIC 581 of the Junta de Andalucía.

References

1. Alimonti, J.B., Ball, T.B., Fowke, K.R. Mechanisms of CD4+ T lymphocyte cell death in human immunodeficiency virus infection and AIDS. *Journal of General Virology*, **84** (2003), 1649–1661.
2. Bhalla, U.S., Iyengar, R., Emergent properties of networks of biological signaling pathways. *Science*, **283** (1999), 381–387.
3. Blossey, R., Cardelli, L., Phillips, A. A compositional approach to the stochastic dynamics of gene networks. *Transactions on Computational Systems Biology*, IV, Lecture Notes in Computer Science, **3939** (2006), 99–122.
4. Cheng, E.H., Wei, M.C., Weiler, S., Flavell, R.A., Mak, T.W., Lindsten, T., Korsmeyer, S.J. BCL-2, BCL-XL sequester BH3 domain-only molecules preventing BAX- and BAK-mediated mitochondrial apoptosis. *Molecular Cell*, **8** (2001), 705–711.
5. Cheruku, S., A. Păun, F.J. Romero, M.J. Pérez–Jiménez, O.H. Ibarra. Simulating FAS–induced apoptosis by using P systems. *Proceedings of the First International Conference on Bio–Inspired Computing: Theory and Applications*, Wuhan, China, September, 18–22, 2006.
6. Hua, F., Cornejo, M., Cardone, M., Stokes, C., Lauffenburger, D. Effects of Bcl-2 levels on FAS signaling-induced caspase-3 activation: Molecular genetic tests of computational model predictions. *The Journal of Immunology*, **175**, 2 (2005), 985–995 and correction **175**, 9 (2005), 6235–6237.
7. Ibarra, O.H., Păun, A. Counting time in computing with cells. *Proceedings of DNA Based Computing, DNA11*, London, Ontario, 25–36, 2005.
8. Kerr, J.F., Wyllie, A.H., Currie, A.R. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. *British Journal Cancer*, **26** (1972), 239.
9. Pogson, M., Smallwood, R., Qwarnstrom, E., Holcombe, Formal agent–based of intracellular chemical interactions. *BioSystems*, **85**, 1 (2006), 37–45.
10. Holcombe, M., Gheorghe, M., Talbot, N. A hybrid machine model of rice blast fungus, Magnaphorte Grisea. *BioSystems*, **68**, 2–3 (2003), 223–228.
11. Jaatela, M. Multiple cell death pathways as regulators of tumour initiation and progression. *Oncogene*, **23** (2004), 2746–2756.
12. Jackson, D., Holcombe, M., Ratnieks, F. Trail geometry gives polarity to ant foraging networks. *Nature* **432** (2004), 907–909.
13. Moehren G., Markevich, N., Demin, O., Kiyatkin, A., Goryanin, I., Hoek, J.B., Kholodenko, B.N. Temperature dependence of the epidermal growth factor receptor signaling network can be accounted for by a kinetic model, *Biochemistry* **41** (2002), 306–320.
14. Moghal, N., Sternberg, P.W. Multiple positive and negative regulators of signaling by the EGFR receptor. *Curr. Opin. Cell Biology*, **11** (1999), 190–196.
15. Nijhawan, D., Honarpour, N., Wang, X. Apotosis in neural development and disease. *Annual Reviews Neuroscience*, **23** (2000), 73–87.
16. Oltavi, Z.N., Milliman, C.L., Korsmeyer, S.J. Bcl-2 heterodimerizes in vivo with a conserved homolog, Bax, that accelerates programmed cell death. *Cell*, **74**, 4 (1993), 609–619.
17. Gh. Păun, Computing with membranes. *Journal of Computer and System Sciences*, **61**, 1 (2000), 108–143, and *Turku Center for Computer Science-TUCS Report* Nr. 208, 1998.
18. Gh. Păun, *Membrane Computing. An Introduction*. Springer-Verlag, Berlin, 2002.

19. Gh. Păun, G. Rozenberg, A guide to membrane computing. *Theoretical Computer Science*, **287** (2002), 73–100.
20. Pérez-Jiménez, M.J., Romero-Campero, F.J. P systems, a new computationl modelling tool for systems biology, *Transactions on Computational Systems Biology VI, LNBI* **4220**, 2006, 176–197.
21. Pérez-Jiménez, M.J., Romero-Campero, F.J. A study of the robustness of the EGFR signalling cascade using continuous membrane systems. *Lecture Notes in Computer Science*, **3561** (2005), 268 – 278.
22. Regev, A., Shapiro, E. (2004) The π-calculus as an abstraction for biomolecular systems. In G. Ciobanu and G. Rozenberg, editors, *Modelling in Molecular Biology*, Springer Berlin.
23. Scaffidi, C., Fulda. S., Srinivasan, A., Friesen, C., Li, F., Tomaselli, K.J., Debatin, K.M., Krammer, P.H., Peter, M.E. Two CD95 (APO-1/Fas) signaling pathways. *The Embo Journal*, **17** (1998), 1675–1687.
24. Schoeberl, B., Eichler–Jonsson, C., Gilles, E.D., Muller, G. Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors, *Nature Biotechnology*, **20**, 4 (2002), 370–375.
25. Van Kampen, N.G. *Stochastics Processes in Physics and Chemistry*, Elsevier Science B.V., Amsterdam, 1992.
26. Walker, D.C., Southgate, J., Hill, G., Holcombe, M., Hose, D.R., Wood S.M., MacNeil, S., Smallwood, R.H. The epitheliome: modelling the social behaviour of cells. *BioSystems*, **76**, 1–3 (2004), 89–100.
27. Wang, K., Yin, X.M., Chao, D.T., Milliman, C.L., Korsmeyer, S.J. BID: a novel BH3 domain-only death agonist. *Genes & Development*, **10** (1996), 2859–2869.
28. Wells, A. EGFR–receptor. *Int. Journal Biochem. Cell Biology*, **31** (1999), 637–643.
29. Wiley, H.S., Shvartsman, S.Y., Lauffenburger, D.A. Computational modeling of the EGFR–receptor system: A paradigm for systems biology. *Trends in Cell Biology*, **13**, 1 (2003), 43–50.
30. ISI web page: `http://esi-topics.com/erf/october2003.html`
31. SciLab Web Site `http://scilabsoft.inria.fr/`
32. P Systems Modelling Framework Web Site: `http://www.dcs.shef.ac.uk/~marian/PSimulatorWeb/P_Systems_applications.htm`