A Study of the Robustness of the EGFR Signalling Cascade Using Continuous Membrane Systems

Research Group on Natural Computing,
Department of Computer Science and Artificial Intelligence,
University of Sevilla,
Avda. Reina Mercedes s/n, 41012, Sevilla, Spain
{marper, fran}@us.es

Abstract. Many approaches to anticancer treatment have had a limited success. A fundamental hurdle to cancer therapy is the robustness of the signalling networks involved in tumourgenesis. The complexity of networks of biological signalling pathways is such that the development of simplifying models is essential in trying to understand the wide-ranging cellular responses they can generate. In this paper a model of the epidermal growth factor receptor signalling cascade is developed using continuous membrane systems. This model is used to study the robustness of this signalling cascade which is known to play a key role in tumour cell proliferation, angiogenesis and metastasis.

Keywords: membrane computing, EGFR signalling network, signal transduction, robustness.

1 Introduction

Membrane Computing is an emergent branch of Natural Computing introduced by Gh. Păun in [9]. Since then it has received important attention from the scientific community. In fact in 2003 the Institute for Scientific Information (ISI) has considered the seminal paper [9] as *fast breaking* and Membrane Computing has been selected as a *fast emerging area* in computer science.

This new non-deterministic model of computation starts from the assumption that the processes taking place in the compartmental structure of a living cell can be interpreted as computations. The devices of this model are called *P systems*. Roughly speaking, a P system consists of a cell-like membrane structure, in the compartments of which one places multisets of objects which evolve according to given rules in a synchronous non-deterministic maximally parallel manner.

Most variants of membrane systems have been proved to be computationally complete, that is equivalent in power to Turing machines, and computationally efficient, that is being able to solve computationally hard problems in polynomial time trading time for space. P systems as a discrete model of computation have

also been used to model biological phenomena (see the volume [1]), and as a continuous model in [7]. A first formalization of non-discrete P system and a way to approximate them was introduced in [2].

In this paper we use a continuous variant of P systems, different from that in [7], to model the epidermal growth factor receptor (EGFR) signalling cascade. Up to now the usual mathematical formalization of biochemical signalling networks has been done using differential equations which are focused on the description of the change in concentration of chemical compounds. Here we use a new formalization of these phenomena in a computational framework which focuses on the compartmental structure (membrane structure) of the cell and on the chemical reactions (rules) that take place in different regions of the cell. Thus this new approach makes possible a topological and modular modelling of intracellular signalling networks. In this framework expansion of an existing model is done by adding new rules (reactions) and, if it is necessary, new membranes to represent new regions (organella) of the cell; therefore the previous model does not need to be changed. Besides modularity and easy extensibility, in favour of our approach we also mention the easy understandability and programmability, features which are not easily achieved in models which use differential equations.

The epidermal growth factor receptor (EGFR) belongs to the tyrosine kinase family of receptors. 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 and degradation in endosomals. Two principal pathways lead to activation of Ras-GTP by hydrolization of Ras-GDP. One of these pathways depends on the concentration of the Src homology and collagen domain protein (Shc) and the other one is Shc-independent. Ras-GTP acts like a *switch* that 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. Disregulated EGFR expression, ligand production and signalling have been proved to have a strong association with tumourgenesis. As a result of this, EGFR has been identified as a key biological target for the development of novel anticancer therapies.

The paper is organised as follows. Continuous P systems are introduced in the next section. In section 3 the EGFR signalling cascade is briefly described. We present our model using continuous P systems in section 4. Results and discussion are exposed in section 5. Finally, conclusions and future work are given in the last section.

2 Continuous P Systems

Usual variants of P systems are discrete models of computation where in every step the rules are applied in a maximal way an integer number of times, we refer to [10] for details. Here we use a variant whose systems can evolve in every instant applying a maximal set of rules a positive *real number of times*

determined by a certain function $\mathcal{K}$. This variant is inspired by the fact that in vivo chemical reactions evolve in a continuous way following a *rate* that depends on the concentration of the reactants.

Roughly speaking, a continuous P system consists of a membrane structure, a hierarchically arranged set of membranes, where one places multisets of objects that represent the concentration of chemical substances. Usual P systems deal with discrete multisets over an alphabet Σ but here we work with continuous multisets (mappings from Σ to $\mathbf{R}^+$, the set of non-negative real numbers). These multisets evolve according to a finite set of rules that represent chemical reactions.

Next we give a formal definition of continuous P systems. A *continuous* P *system* is a construct, $\mathbf{\Pi} = (\Sigma, \mu, w_1, \ldots, w_n, \mathcal{R}, \mathcal{K})$, where:

1. $n \geq 1$ is the degree of the system (number of membranes).
2. $\Sigma = \{c_1, \ldots, c_m\}$ is the alphabet of *objects*.
3. μ is a *membrane structure* (a rooted tree) consisting of n membranes (nodes of the tree) labelled with $1, \ldots, n$ (often, we identify the membranes with labels from a finite set H).
4. $w_1, \ldots, w_n$ are continuous multisets associated with each membrane of the membrane structure μ.
5. $\mathcal{R}$ is a finite set of *rules* of the form:

$$u\ [\ v\ ]_i\ \rightarrow\ u'[\ v'\ ]_i,$$

   where $u, v \in \Sigma^*$ represent the reactants, $u', v' \in \Sigma^*$ represent the products, and $i \in H$ is the label of the relevant membrane of the reaction that is modelled.
6. $\mathcal{K}$ is the *rate of application function* which associates with each rule and multiplicity of the objects in μ, a non-negative real number considered as the rate of application of the rule:

$$\mathcal{K} : \mathcal{R} \times \mathcal{M}_{n\times m}(\mathbf{R}^+) \rightarrow \mathbf{R}^+,$$

   where $\mathcal{M}_{n\times m}(\mathbf{R}^+)$ is the set of matrices of order $n \times m$ over $\mathbf{R}^+$.

A *configuration* of a continuous P system $\mathbf{\Pi}$ is a matrix of $\mathcal{M}_{n\times m}(\mathbf{R}^+)$ where the object in row i and column j, $a_{i,j}$, represents the multiplicity of the object c_j in the membrane i. We interpret the configurations as assignments of continuous multisets to the membranes of the system, that is, the association of each region with the concentration of chemical substances present in it.

For usual P systems we talk about *computations* but for continuous P systems we prefer to think of *evolutions*. An *evolution* of a continuous P system is a mapping from $\mathbf{R}^+$ to $\mathcal{M}_{n\times m}(\mathbf{R}^+)$. That is, an evolution E associates with each instant $t \in \mathbf{R}^+$ an instantaneous configuration $E(t)$ of the system:

$$E(t) = (a_{i,j}(t))_{\substack{1 \leq i \leq n \\ 1 \leq j \leq m}}$$

For each $t \in \mathbf{R}^+$ and i, $1 \leq i \leq n$, we denote by $v_i(t)$ the continuous multisets over $\Sigma = \{c_1, \ldots, c_m\}$ defined as follows: $(v_i(t))(c_j) = a_{ij}(t)$ for $1 \leq j \leq m$. That is, we can describe $E(t)$ by a tuple $(v_1(t), \ldots, v_n(t))$.

The way a continuous P system, $\mathbf{\Pi} = (\Sigma, \mu, w_1, \ldots, w_n, \mathcal{R}, \mathcal{K})$, evolves is determined by the initial multisets $w_1, \ldots, w_n$ and the rate of application function $\mathcal{K}$. We define the *initial configuration* of $\mathbf{\Pi}$ as the tuple $(w_1, \ldots, w_n)$.

The rules are applied during the evolution of the system in a continuous way according to the *rate of application function* $\mathcal{K}$. At an instant $t \in \mathbf{R}^+$, a rule $r \in \mathcal{R}$ is applied exactly $\mathcal{K}(r, E(t))$ *times* (in this sense, we can say that the rules are applied in a $\mathcal{K}$-maximal way); that is, $\mathcal{K}(r, E(t))$ units of the reactants are consumed and $\mathcal{K}(r, E(t))$ units of the products are produced. Observe that the effect of the rule r decreases the multiplicity (concentration) of its reactants and increases the multiplicity (concentration) of its products. More precisely, we define the effect of a rule r during an interval of time $[t, T]$ as follows:

$$Ef(r, t, T) = \int_t^T \mathcal{K}(r, E(s))\, ds.$$

More formally, given an object $c_j \in \Sigma$, $1 \leq j \leq m$, and a membrane i, $1 \leq i \leq n$, we denote by $production_i(c_j)$ (resp. $consumption_i(c_j)$) the set of rules where c_j is a product in membrane i (resp. a reactant). Therefore the real number $(v_i(t))(c_j)$, denoted by $|c_j|_i(t)$, is determined by the next formula:

$$|c_j|_i(t) = |c_j|_i(0) + \sum_{r \in production_i(c_j)} \int_0^t \mathcal{K}(r, E(s))\, ds - \\ - \sum_{r \in consumption_i(c_j)} \int_0^t \mathcal{K}(r, E(s))\, ds,$$

with $v_i(0) = w_i$, that is, $|c_j|_i(0) = w_i(c_j)$).

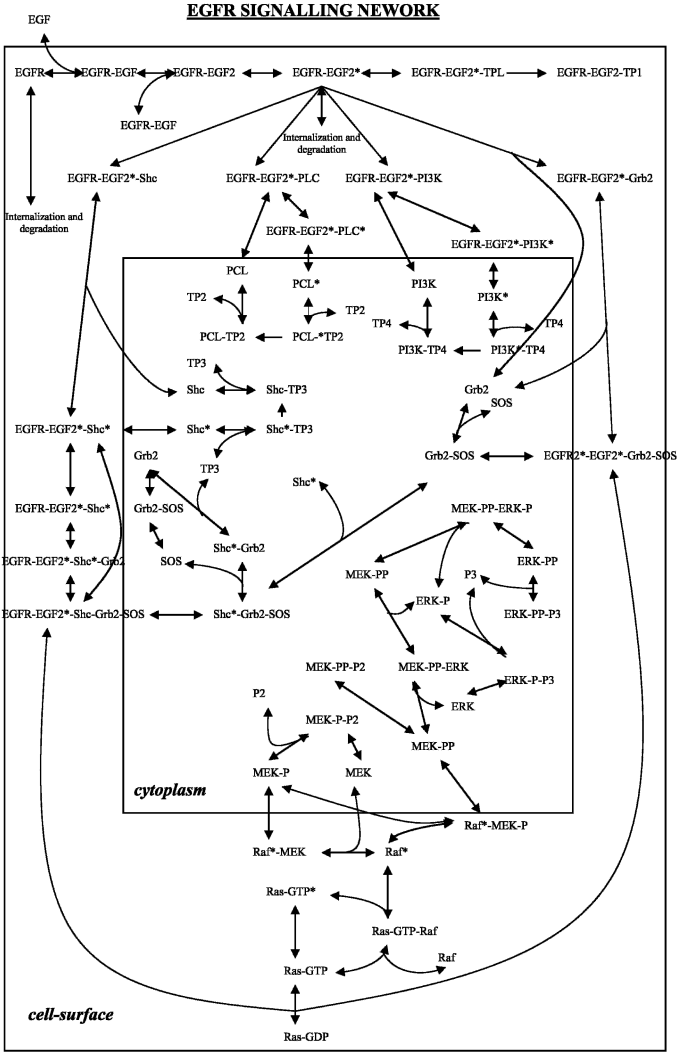
In computers, real numbers are represented by a finite set of rational numbers. Therefore, like in most continuous models we need to develop approximations in order to simulate evolutions of continuous P systems in computers.

As shown above, in order to determine the effect of a rule on the evolution of a system during an interval of time $[t, T]$ we only need to compute an integral of the rate of application function $\mathcal{K}$. Hence, in order to approximate the evolution of a continuous P systems in a finite set of instants $t_0, \cdots, t_q$ we can use any suitable known numerical method to approximate integrals. Here for simplicity we use the rectangle rule; that is, we suppose $t_{l+1} - t_l = p$ is *small enough* to assume that $\mathcal{K}$ remains constant and equal to $\mathcal{K}(r, E(t_l))$ in the interval $[t_l, t_{l+1}]$ for $l = 0, \ldots, q-1$. With this assumption we can approximate the effect of a rule during an interval of time of length p by $Ef(r, t_l, t_{l+1}) \approx p\mathcal{K}(r, E(t_l))$.

By doing this approximation we reach an usual P systems that performs q steps $(t_0, \ldots, t_q)$ and in each steps the rules are applied $p\mathcal{K}(r, E(t_l))$ times. Therefore, we have approximated the evolution of a continuous P system by the computation of an usual discrete P system working in a $p\mathcal{K}$ bounded parallel manner.

3 EGFR Signalling Cascade

In this section we will describe briefly the EGFR signalling cascade following the network depicted below. During the signal transduction which takes place in this cascade, the information about the concentration of the EGF in the outside of the cell is translated into kinetic information inside the cell by EGFR phosphorylation.



The epidermal growth factor receptor (EGFR) belongs to the tyrosine kinase family of receptors. The binding of the epidermal growth factor (EGF) to the

extracellular domain of EGFR induces receptor dimerization and autophosphorylation of intracellular domains. Then, on the one hand, a multitude of proteins are recruited starting a complex signalling cascade and, on the other hand, the receptor follows a process of internalization, ubiquitination and degradation.

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 consist 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 suggest a strong robustness of the system.

For a more detailed description of the cascade see the literature listed in the bibliography.

4 Modelling EGFR Signalling Cascade by Continuous P Systems

We have developed a model of the signalling cascade described in the previous section using a continuous P system, $\mathbf{\Pi}_{EGF} = (\Sigma, \mu, w_e, w_s, w_c, \mathcal{R}, \mathcal{K})$. Our model consists of more that 60 proteins and complexes of proteins and 160 chemical reactions. *Supplementary information and details* about the model are available on the web page `www.gcn.us.es/egfr.pdf`.

• **Alphabet:** In the alphabet Σ we collect all the proteins and complexes of proteins that take part in the signalling cascade. In table 1 of the *Supplementary*

information all the objects of the alphabet and the chemical compounds that they represent are listed.

• **Membrane Structure:** In the EGFR signalling cascade described in the previous section, there are three relevant regions, namely the *environment*, the *cell surface* and the *cytoplasm*. We represent them in the membrane structure as the membranes labelled with: e for the environment, s for the cell surface and c for the cytoplasm.

• **Initial Multisets:** In the initial multisets we represent the initial concentrations of the chemical substances in the environment, the cell surface and the cytoplasm. These concentration has been obtained from the references listed in the bibliography. A detailed presentation of initial multisets is shown in table 2 of the *Supplementary information*.

• **Rules and Rate of application function:** In the rules we model the chemical reactions described which form the signalling cascade. To model the reactions we use the *Law of Mass Action* which states that the rate of a reaction is proportional to the product of the concentrations of the reactants. That is, if we have a reaction of the form:

$$r_1 + \cdots + r_k \rightarrow p_1 + \cdots + p_{k'},$$

then the rate of this reaction is $k|r_1| \cdots |r_n|$, where k is called *kinetic constant*.

In tables 3-11 of the *Supplementary information* (`www.gcn.us.es/egfr.pdf`) all the rules are listed as well as the kinetic constants and the references from where they were taken. As an example of the procedure we have followed to develop our model, we next present the derivation of one of the 160 rules.

Let us consider the binding of EGF to EGFR:

$$\text{EGF EGFR} \rightarrow \text{EGF-EGFR}$$

We know from biological experiments that EGF, which is present in the environment, binds to EGFR, which is present in the cell surface at a rate of $0.003\, nM^{-1}s^{-1}$. According to this, the relevant membrane in this reaction is the cell-surface because it separates the two regions involved in this reaction. Besides following the Mass Action Law the reaction takes place at a velocity of $0.003|EGF||EGFR|$. Therefore, in our model we represent this chemical reaction by the following rule and rate of application:

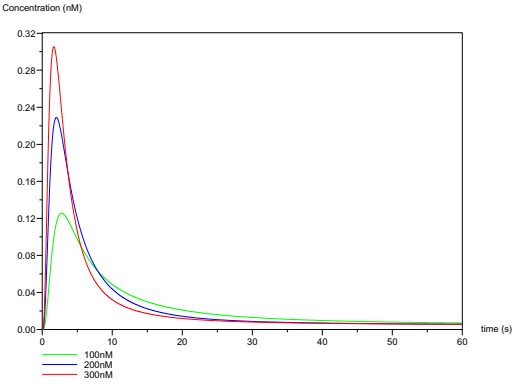
$$\text{EGF } [\text{EGFR}]_s \rightarrow [\text{EGF-EGFR}]_s \quad \mathcal{K}(r, E(t)) = 0.003|EGF(t)|_e|EGFR(t)|_s$$

5 Results and Discussion

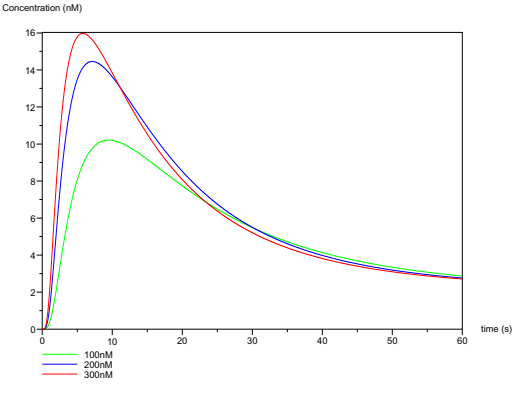
The model presented in the previous section has been implementing using CLIPS, a productive development and delivery expert system tool which provides a complete environment for the construction of rule and/or object based expert systems.

To implement our model we have approximated the evolutions of the continuous P system $\mathbf{\Pi}_{EGF}$ by computations of an usual P system working in a bounded parallel manner. The parameter p, chosen for the approximation, was fixed to 10^{-3} after testing different values until the results obtained did not change.

We study the effect of different EGF concentrations on the signalling cascade. To illustrate this effect we depict the evolution of the concentration of the most relevant proteins in the signalling cascade over time.



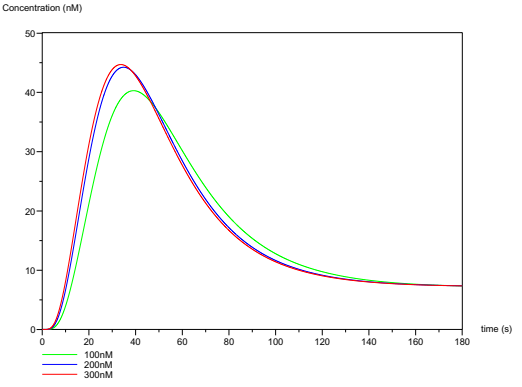
Above it can be seen that the receptor activation by autophophorylation is clearly concentration dependent showing a high peak in the first 5 seconds to decay rapidly afterwards to very low levels of concentration. According to the variance in the receptor activation it is intuitive to expect different cell responses to different EGF concentrations. Here we will show that this is not the case. Next we show the evolution of the phosphorylation of Shc after binding to the receptor over 60 seconds.



It can be observed that the responses to EGF stimulation get more sustained as we get deeper in the cascade and that the network has almost managed to

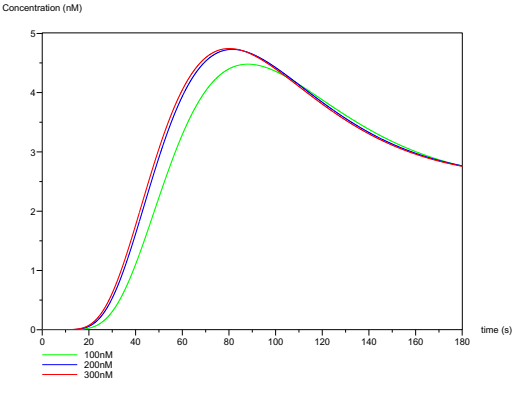
attenuate the overstimulation with EGF. Observe that the red and blue line are almost identical.

The activation of the Ras protein is a key node in the EGFR signalling cascade. We depict its evolution over 180 seconds in the next graphic.



Note that at this point the response to the overstimulation with EGF has been completely attenuated and that an amplification of the response to the low concentration has been performed.

Finally, the goal of the cascade is the activation of the mitogenic kinases MEK and ERK. These proteins regulates the transcription of several proto-oncogenes like c-fos and c-Jun. Next the evolution of phosphorylated MEK over 180 second is presented.



In this picture it is shown the surprising robustness of the signalling cascade. The signals from outside due to EGF concentration have been either attenuated or amplified to get the same concentration of the most relevant kinases. Observe that after 100 seconds, when the response gets sustained, the three line representing the response to different external EGF concentrations are identical. For more

graphics of the evolution of the different proteins in the dynamics of the cascade see the *Supplemtary information* on the web page `www.gcn.us.es/egfr.pdf`.

Currently, the robustness of the EGFR signalling cascade is proposed to be a product of receptor internalization and cross-talk between different pathways in the cascade. According to the literature listed in the bibliography receptor internalization produces a signal attenuation by protection from high external EGF concentration, meanwhile an amplification of the signal due to low EGF concentrations is performed in several nodes of the cascade where there exists a cross-talk between different pathways of the cascade. Our model is in accordance with this hypothesis. This outcome shows the reliability of our model to make post-diction and supports the possibility of using our model to produce new hypotheses and predictions about the behaviour of this relevant network in the cell cycle and tumourgenesis.

6 Conclusions and Future Work

In this paper we have developed a topological and modular model of the EGFR signalling cascade which consists in more than 60 proteins and complexes of proteins and 160 chemical reactions. Our model can provide insight into the dynamics of the MAP kinase cascade which is activated by EGFR autophosphorylation. It can be also useful to formulate hypotheses that can be tested experimentally. Actually, our model suggests that the the cascade is robust to variation in the EGF concentration; therefore in the next future we intend to study the influence of kinase inhibition at different cytoplasmic nodes of the signalling cascade.

The results obtained using this model are in well agreement with experimental data. This shows that continuous membrane systems are a reliable framework for modelling networks of biochemical signalling cascades.

Currently our model is being translated into *SBML* (System Biology Markup Language) a computer-readable format like XML for representing models of biochemical reaction networks. Moreover an user friendly interface for the CLIPS implementation is being designed using JAVA. We hope that these two current works will help to spread our model in the scientific community and so the authors, who are not biologists, can get some feedback from specialists in networks of biochemical signalling cascades.

Finally, this model takes a first step (PI3K phosphorylation) towards the MDM2-p53 feedback loop; we intend to expand our model to comprise this interaction between proteins which it is known to play a key role in the regulation of cell cycle and tumourgenesis.

Acknowledgement

This work is supported by Ministerio de Ciencia y Tecnología of Spain, by *Plan Nacional de I+D+I (2000–2003)* (TIC2002-04220-C03-01), cofinanced by FEDER funds, and by a FPI fellowship from the University of Seville.

References

1. Ciobanu, G.; Păun, Gh.; Pérez-Jiménez, M.J., eds: *Applications of Membrane Computing*, Springer-Verlag (2005) in press.
2. Cordón-Franco, A.; Sancho-Caparrini, F.: Approximating Non-Discrete P Systems. In G. Mauri, Gh. Păun, M.J. Pérez-Jiménez, Gr. Rozenberg, A. Salomaa (eds.) *Membrane Computing, International Workshop, WMC5, Milan, Italy, June 2005, Revised Papers*. Lecture Notes in Computer Science, Springer, **3365**, in press.
3. El-Masri, H.A.; Portier, C.J.: Replication Potential of Cells via the Protein Kinase C-MAPK pathway: Application of a Mathematical Model, *Bulletin of Mathematical Biology* **61** (1999) 379–398.
4. Kholodenko, B.G. et al.: Quantification of Short Term Signalling by the Epidermal Growth Factor Receptor. *J. Biol. Chem.*, **274** (1999) 30169–30181.
5. Miller, J.H.; Zheng, F.: Large-scale Simulations of Cellular Signaling Processes, *Parallel Computing* **30** (2004) 1137–1149.
6. Moehren G. et al: Temperature Dependence of the Epidermal Growth Factor Receptor Signaling Network Can Be Accounted for by a Kinetic Model, *Biochemistry* **41** (2002) 306–320.
7. Nishida, T.Y.: Simulations of Photosynthesis by a K-Subset Transforming System with Membranes, *Fundamenta Informaticae*, **49**, 1 (2002), 101–113.
8. Păun, A.; Păun, Gh.: The Power of Communication: P Systems with Symport/Antiport, *New Generation Computing*, **20**, 3 (2002), 295–305.
9. Păun, Gh.: Computing with Membranes, *Journal of Computer and System Sciences*, **61**(1) (2000) 108 – 143.
10. Păun, Gh.: *Membrane Computing. An Introduction*, Springer-Verlag Berlin, 2002.
11. Pérez-Jiménez, M.J.; Romero-Campero, F.J.: Modelling EGFR Signalling Network using Continuous Membrane Systems, Accepted paper in the Workshop on Computational Methods in Systems Biology 2005
12. Plank, M.J.; Sleeman, B.D.: Tumour-Induced Angiogenesis: A Review, *Journal of Theoretical Medicine*, in press.
13. Schoeberl, B. et al.: Computatinal Modeling of the Dynamics of the MAP Kinase Cascade Activated by Surface and Internalized EGF Receptors, *Nature Biotech.* **20** (2002) 370–375.
14. Sermon, B.A. et al: The Importance of Two Conserved Arginine Residues for Catalysis by the Ras GTPas-activating Protein, Neurofibromin, *The Journal of Biological Chemistry* **273** (1998) 9480–9485.
15. Starbuck, C.; Lauffenburger, D.A.: Mathematical Model for the Effects of Epidermal Growth Factor Receptor Trafficking Dynamics on Fibroblast Proliferation Responses, *Biotechnol. Prog.* **8** (1992) 132–143.
16. Sydor, J.R. et al: Transient Kinetic Studies on the Interaction of Ras and the Ras-Binding Domain of c-Raf-1 Reveal Rapid Equilibration of the Complex, *Biochemistry* **37** (1998) 14292–14299.
17. Wiley, H.S.: Anomalous Binding of Epidermal Growth Factor to A431 Cells is Due to the Effect of High Receptor Densities and a Saturable Endocytic System, *The Journal of Cell Biology* **107** (1988) 801–810.
18. ISI web page `http://esi-topics.com/erf/october2003.html`
19. P systems web page `http://psystems.disco.unimib.it/`