Topics and Problems in Metabolic P Systems

Vincenzo Manca

University of Verona
Department of Computer Science
strada Le Grazie, 15
37134 Verona, Italy
vincenzo.manca@univr.it

Summary. P metabolic systems are a special class of P systems which seem to be adequate for expressing biological phenomena related to metabolism and signaling transduction in biological systems. We give the basic motivation for their introduction and some ideas about their applicability and development.

1 Introduction

P systems were introduced in [13] as a new computation model inspired by biology. The two main aspects of a P system are *multisets* and *membranes*. The notion of multiset is related to the way sets are *implemented* in terms of physical entities. In fact, any object can occur in a certain number of instances, that is, different *concrete* individuals (*objects, occurrences, copies*), located in space and time, which can be seen as expressing the same *abstract* individual identity, or shortly the same *type*. The different features of concrete individuals are regarded as inessential with respect to some pertinent traits specific of their abstract identity. Only a finite number of copies can be assumed for any individual, at any time, because each copy possesses a mass, and the overall mass of any physical system is assumed to be bounded. In physical reality, any type is always observed by means of some concrete individual that represent some (abstract) identity among a number of possible observable (abstract) identities. This point of view fits completely with the notion of molecule. When we say a Carbon molecule, we just intend to say a physical entity having the chemical type of Carbon. But, in chemical and biochemical systems, in almost all cases, we deal with a huge number of concrete individuals, that is, populations of objects, where a population is specified by some types and by the number of objects of each type. In biochemical systems, objects (molecules) are localized in different compartments. These compartments play a crucial role for life strategies, mainly: *selection, concentration, protection*. Inside a membrane some objects are selected which are useful for maintaining some reactions. These objects put in the same spatial vicinity react in a better way than in a unbounded

area. Moreover, the membrane boundary protects from the external agents which can disturb processes inside the membrane, and at same time, provides a filter to control the relationship with the external world. In this perspective, the state of a system is given by the number of objects of each type present at a given instant and by how they are distributed in the membranes of the system.

The theory of P systems has grown very fast [14, 3, 17] by studying different kinds of evolution rules (reactions) and different kinds of evolution strategies. In P systems, the passage from a state to another one is produced by the application of rules which act, independently in each membrane, transforming multisets. For example, a rule, denoted by $AA \to AB$ (acting in a given membrane) can be applied if two objects of type A are present. Therefore, if applicable, this rule is applied and any two objects A are removed and replaced by one A and one B. This means that at least two A must be present in the membrane, but only one of them will be replaced by one B. If more rules are applicable, then a maximal set of rules, applicable in a parallel way, is chosen and then all of them are applied. This strategy is commonly referred as "maximal parallel rewriting" and generally it assumes nondeterministic evolutions (from a given state, transitions to different states are possible). The main aspect investigated in this context has been the computational power of different kinds of P systems.

In a very first approximation, a cell is a membrane system, and moreover, its functioning is determined by all the types of molecules inside it, the amount of copies of these types, and the cell compartments where they are located. Therefore, it is of great importance to define a method for computing the evolution of a P system that is directly meaningful with respect to biological processes. In this perspective, a transformation $AA \to BC$ is better read in chemical terms, as something which expresses the following prescription: "two *moles* of A produce one mole of B and a mole of C", where a *mole* is a population unit. But, when many reaction are working together, a competition among reactions needing the same kinds of reactants is better expressed by the notion of *reaction unit* which is not conceived in an absolute way, as it happens in the classical chemical setting (1 mole $\approx 6.02 \times 10^{23}$ molecules), but it is relative to each rule, to each state of the system and to the rules being in competition. However, a sort of Avogadro's principle has to be kept, that is, a rewriting rule has to be read in a stoichiometric way. For example, a rule such as $AA \to BC$ should reasonably say that $2m$ objects of type A have to be consumed by the rule, and m objects B plus m objects C have to be produced by it. The crucial point of this discussion is "how has the number m to be calculated" in order to reproduce adequately a given biochemical process?" This problem becomes more difficult than it may appear at a first glance: in fact, such a number has to be assigned to any rule (working in a given membrane) and it depends on the current state of the system (the types of objects with their relative quantities).

We addressed this question, in the context of P systems, in a series of papers [1, 10, 4, 2] were several proposals were developed along the same line of thought. In this paper we will outline the most recent answer we found, based on the P

metabolic algorithm, shortly PMA, [11], and we list some situations where PMA provided satisfying models of biological dynamics. Finally, we discuss some important problems that deserve further research and the developments of tools directly usable by biologists for modeling phenomena in an easy, but rigorous language, and for simulating and evaluating them by computational supports.

We summarize the principles of PMA with four statements:

(1) Rules compete for object populations.
(2) Objects are allocated to rules according to a *mass partition principle*.
(3) Partition factors are determined by *reaction maps*.
(4) A "Metabolic rule" r consumes/produces integer multiples of a *reaction unit* u_r which generalizes the notion of molar unit (Avogadro's principle).

It may be useful to clarify these principles by means of an example. Let $T = \{A, B, C, \ldots\}$ be an alphabet of biological species (or types), and define $q : T \to \mathbb{N}$ as the state of the system, that is, the concentration of each type at a certain observation instant. Assuming that at a given instant four rules, say $r_2, r_3, r_5,$ and r_7, need molecules of a certain type A for performing some biochemical reactions (see Figure 1), then a partition strategy for species A is necessary. The novelty of our approach is that of considering a real number as the

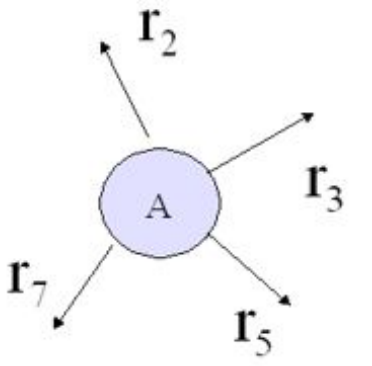

Fig. 1. Competition for object A.

strength of any rule. This real number is the value that the *reaction map* associates to the rule in the considered instant. For example, with respect to Figure 1, if $q(A) = a, q(B) = b, q(C) = c$, then the reactivities associated to the rules $r_2, r_3, r_5,$ and r_7 which ask for A molecules could be:

$$f_2 = 200a, f_3 = 0.5a^{1.25}b^{-1}, f_5 = a^{1.25}(b+c)^{-1}, \text{ and } f_7 = 10.$$

We define

$$K_{A,q} = \sum_{i=2,3,5,7} f_i(q)$$

as the *total pressure* on A which, for simplification, we assume to be always a positive value. Then, for each of the rules r_j we consider the *partial pressure* (or *weight*) of r_j on type A as

$$w_{A,q}(r_j) = \frac{f_j(q)}{K_{A,q}} \ .$$

Getting back to the example discussed before, it should be easy to see that

$$w_{A,q}(r_2) = \frac{200a}{200a + 0.5a^{1.25}b^{-1} + a^{1.25}(b+c)^{-1} + 10}$$

while

$$w_{A,q}(r_3) = \frac{0.5a^{1.25}b^{-1}}{200a + 0.5a^{1.25}b^{-1} + a^{1.25}(b+c)^{-1} + 10}$$

and the other weights can be calculated analogously. These weights determine the partition factor of the species A in the state q.

Let us assume that, at a given instant, according to the reaction competition, n objects of type A and m objects of type B were allocated to a rule $r : AAB \to AC$. The corresponding reaction unit turns out to be

$$u_r = min\{n/2, m\}$$

and this means that $2u_r$ objects of type A and u_r objects of type B are consumed, while u_r objects of type A and C are produced. This globally states that u_r objects of type A and u_r objects of type B are replaced by u_r objects of type C. But, the important thing to point out here is that rule r is absolutely different from a rule r' having the form $AB \to C$, and this is due to the fact that the two rules imply different competition factors, and consequently, different mass partitions. In fact, in the second case the reaction unit would have been $u_{r'} = min\{n, m\}$.

2 Intermezzo

Before going into the details of P metabolic algorithm, it could useful to address two topics: one more specific, the other one more general.

In our search for methods which compute the evolution of P metabolic systems, we implicitly assume that these systems are deterministic. This aspect distinguishes our systems from P systems in their generality, which are intrinsically nondeterministic. However, it is important to realize that the determinism we consider is a special type of determinism we could say a "population determinism". In fact, we do not intend to predict what happens to the single objects, but only to evaluate the distribution of the objects at a given observation instant, when their distribution at the previous instant was given. This means that at a macroscopic

level a system is determined, because its macroscopic appearance depends only on the biochemical types distribution, but many different *microstates* are associated to the same *macrostate*. This is not so different from the physical state of a gas, given by its volume, pressure and temperature, but corresponding to a huge number of micro-mechanical states of its molecules, and where the relationship between these two physical levels has a statistical nature (the velocity distribution among molecule population). The thermodynamic laws allow us to predict the evolution of system, in terms of macroscopic state variables but do not say exactly anything about the single velocities of particles in the system.

The second remark concerns a new perspective arising when algorithmic and computational tools are used in modeling natural phenomena. In 1986, in the inaugurations of the *Laboratory for Foundation of Computer Science* at the Department of Computer Science of the University of Edinburgh, Robin Milner put as title of his lecture [12] the following question: "Is Computing an Experimental Science?" Now, to the extent we are able to afford and to solve (even partially) the problem of predicting the evolution of a biological system by a suitable algorithm (with a good accordance with the experimental data), especially in those cases where classical mathematical tools are not applicable, just to that extent, the answer to that question is definitively positive. The relevance of this remark has a very strong scientific and philosophical meaning. In fact, recently, in a call for PhD students, just from the *School of Informatics* at the University of Edinburgh "Informatics" has been defined as *the discipline which studies information and computation in natural and artificial systems*. In this sense, "Natural Computing" is a field which points out the deep scientific role of "informational methods" as a *clavis* for nature comprehension, comparable to other experimental scientific disciplines, based on mathematics, but with a new and specific emphasis on algorithms and mathematical discrete methods.

3 P Metabolic Algorithm

A *metabolic P system of level 0* (with only the skin membrane), shortly a MP system, is given by a structure $M = (Q, T, R, F, q_0)$, where

- T is an alphabet of *types* of M;
- Q are the *states* of M, which are functions from T to the set $\mathbb{N}$ of natural numbers;
- R is the set of rules of M which are denoted by $\alpha_r \to \beta_r$ with α_r, β_r strings over T, for any $r \in R$;
- $F = \{f_r \mid r \in R\}$ is the set of *reaction maps* of M, which are functions $f_r : Q \to \mathbb{R}$ taking values in the set $\mathbb{R}$ of real numbers;
- $q_0 \in Q$ is the *initial state* of M.

The evolution of M in time is given by a *dynamical function* $\varphi : \mathbb{N} \to Q$ such that $\varphi(0) = q_0$ and, for any $n \in \mathbb{N}$, $\varphi(n+1)$ is calculated from $\varphi(n)$ by means of the metabolic algorithm MPA.

Let us present the P metabolic algorithm for MP systems. Its extension to metabolic P systems with many membranes, say MPM systems, imposes a notational overhead, but follows the same basic logic. The execution of `PMA(M, q, n)` provides as outputs n steps in the evolution of the metabolic system M starting from a state q. The following notation is used in the formulation of the algorithm.

- Sub_r is the set of types that are substrates of rule r;
- $h_r(A)$ and $g_r(A)$ are the numbers of occurrences of symbol A in α_r; and β_r respectively, and $d_r(A) = g_r(A) - h_r(A)$;
- $R(A)$ is the set of rules where symbol A occurs, and $R_\alpha(A)$ is the set of rules where A occurs as a substrate.
- For a finite set S of numbers $min\ S$ is its minimum and $min\ \emptyset = 0$.

```
PMA(M, q, n)
   begin
      For i = 1..n do
            begin
               1. For each symbol A ∈ T do
                        begin
                           k(q, A) := Σ_{r∈R_α(A)} f_r(q);
               2.          For each r ∈ R_α(A) do
                              begin
                                 i.  w_r(q, A) := f_r(q) / (h_r(A)·k(q,A));
                                 ii. m_r(q, A) := ⌊w_r(q, A) · q(A)⌋;
                              end
                        end
               3. For each symbol A ∈ T do
                        begin
               4.          For each r ∈ R(A) do
                              begin
                                 i.  u_r(q) := min{m_r(q, X)|X ∈ Sub_r};
                                 ii. δ_r(q, A) := u_r(q) · d_r(A);
                                 iii. if Sub_r = ∅
                                             λ_r(q, A) := f_r(q) · d_r(A);
                                      else
                                             λ_r(q, A) := 0;
                                      fi
                              end
               5.          Δ(q, A) := Σ_{r∈R(A)} δ_r(q, A);
                           Λ(q, A) := Σ_{r∈R(A)} λ_r(q, A);
               6.          q'(A) := q(A) + Δ(q, A) + Λ(q, A);
                        end
               7. q := q';
               8. output q;
```

```
        end
end
```

The P metabolic algorithm was proven adequate in many case of biological modeling. So far the phenomena listed below were described in terms of P systems and their evolutions were calculated by PMA:

BZ : Belousov-Zhabotinski Brusselator [1, 2],
LV : Prey-Predator Lotka-Volterra Dynamics [1, 2],
SIR : Susceptible-Infected-Recovered Epidemic [1],
LR : Leucocyte Selective Recruitment in Immune Response [6],
PKC : Protein Kinase C Activation [2],
CR : Circadian Rythms [4],
MC : Mitotic Cycles [11].

In Figure 2 is depicted an important phenomenon of signaling mechanism, under investigation by means of PMA, which occurs in *Dictyostelium discoideum* (Dd), an amoeba which is one of the most studied organisms in developmental biology. Dd can switch from unicellular to multicellular stages (isolated and collective phases) by means of a chemical mechanism of intercellular communication with a periodic nature, which presents similarities with hormonal communications in higher organisms [8]. Pulses of Adenosine Monophosphate cAMP are generated with a periodicity of 7 minutes. Once multicellular stage is reached, amoebae differentiate into at least two distinct cell types, thus providing a simple model for the study of pattern formation.

Substances involved in this phenomenon are the following.

cAMP = cyclic adenosine monophoshate,
PDEs = phosphodiesterase (extracellular),
CAR1 = cAMP Receptor 1,
ERK2 = Extracellular Regulated Kinase 2,
ACA = Adenilate Cyclase A,
PKA = Protein Kinase A,
RegA = intracellular Phosphodiesterase.

In Figure 3 is reported a diagram showing the evolution of three substances [9]. A problem under investigation is that of determining a suitable metabolic system which provides the behavior given in Figure 3.

4 Open Problems

Classically, a metabolic system is defined by a set of variables $x_1, x_2, \ldots, x_n$ which satisfy a system of ordinary differential equations (ODE) of the following type [15, 16]:

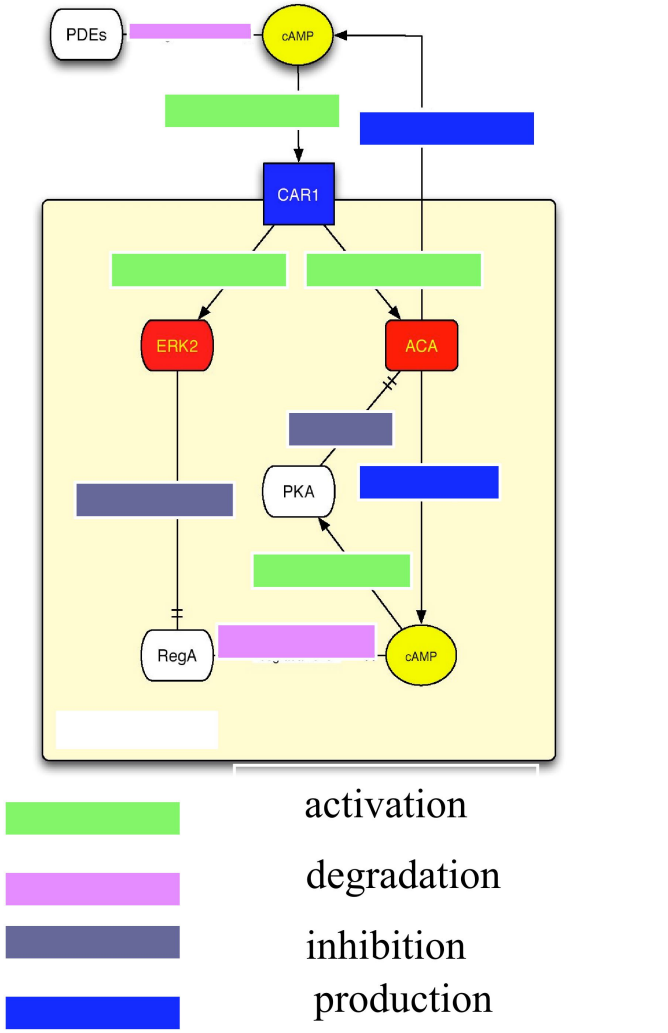

Fig. 2. Dictyostelium discoideum: Chemotactic Signaling Mechanism.

$$
\begin{aligned}
&dx_1/dt = f_1(x_1, x_2, \ldots, x_n),\\
&dx_2/dt = f_2(x_1, x_2, \ldots, x_n),\\
&\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\
&dx_n/dt = f_n(x_1, x_2, \ldots, x_n).
\end{aligned}
$$

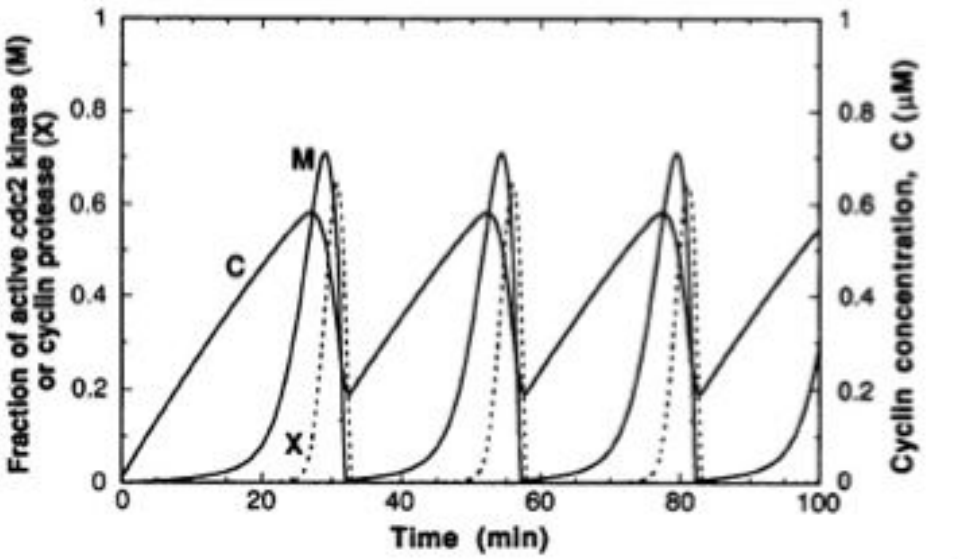

Fig. 3. Dictyostelium discoideum: Evolution of three substances.

Such a system is *autonomous* because the time variable t does not occur explicitly in the right hand members of equations, that is, variable variations depend exclusively on the values at each instant, but no memory is in the states of the system. Now, many natural questions arise: "What is the relationship between the differential formulation of a metabolic system and its formulation in terms of P systems and P metabolic algorithm?" "Do systematic translation methods exist between the two metabolic approaches?" "Which advantages, if any, there are in using PMA instead of ODE (Ordinary Differential Equations), and when these advantages are effective?". In [5] some initial steps in the comparisons between PMA and ODE have been done, and in [11] it is shown a case from [7] relative to the mitotic cycle, in amphibian embryos. Its formulation in P metabolic systems follows directly from the biological analysis of the phenomenon, in a way that is much simpler than the differential equations formulation.

Other two aspects which deserve a further investigation are the search for metabolic systems representing basic oscillators and the implementation of a friendly interface for PMA. The search of metabolic oscillators, and the way they combine in more complex oscillatory patterns, could be very useful in the analysis of complex biological network. In fact, communication and interaction in living systems is very often based on oscillatory mechanisms. Finally, a friendly interface for PMA, with other related tools, could allow biologists to experiment directly the adequacy of the algorithm for a wide class of situations, and therefore, could suggest how to extend and how to improve it (presently, for any system specific rules, reaction maps, and initial states are translated in code lines, therefore any

change, even in a constant, requires to intervene in the code, with a dependence on whom knows the program and with the risk to introduce errors at any simulation).

The strategy of PMA suggests a natural representation of rules as special graphs we called MP graphs which are constituted of two levels: a first level describes the *Stoichiometric Network* of reactions, the second level expresses the *Regulation Network*, which tunes the relative strengths of rules. In [11] this formalism has been described and it is shown how an MP graph can be deduced which provides all the data for computing, via PMA, the evolution of an MP system with a given initial state. This aspect makes PMA a natural tool for analysis and computational evaluation of biological models, directly usable by biologists who can "drive the car" without the need of an expert assistance translating their statements into suitable equations.

At present, there is no clear definition of "Systems Biology", but surely most people would agree it has something to do with understanding dynamic and molecular-level relationship among biological molecules in living systems. For this reason, tools which provide intuitive means for representing and analyzing dynamics of complex biological networks seem us to be a necessary step in the assessment of a discipline that seeks "to connect the dots between molecular data".

References

1. L. Bianco, F. Fontana, G. Franco, V. Manca: P systems for biological dynamics. In[3], 81–126.
2. L. Bianco, F. Fontana, V. Manca: P systems with reaction maps. *International J. of Fundations of Computer Science*, 17, 1 (2006), 27–48.
3. G. Ciobanu, Gh. Păun, M. J. Pérez-Jiménez, eds.: *Applications of Membrane Computing*. Springer, Berlin, 2006.
4. F. Fontana, L. Bianco, V. Manca: P systems and the modeling of biochemical oscillations. In *Membrane Computing, 6th International Workshop, WMC 2005* (R. Freund, Gh. Păun, G. Rozenberg, A. Salomaa, eds.), LNCS 3850, Springer, 2006, 200–209.
5. F. Fontana, V. Manca: Prey-predator dynamics in P systems. Submitted, 2006.
6. G. Franco, V. Manca: A membrane system for the leukocyte selective recruitment. In *Membrane Computing, WMC 2004*, LNCS 2933, Springer, 2004, 181–190.
7. A Goldbeter: A minimal cascade model for the mitotic oscillator involving Cyclin and cdc2 kinase. *PNAS*, 88, 20 (1991), 9107–9111.
8. A. Goldbeter: *Biochemical Oscillations and Cellular Rhythms. The Molecular Bases of Periodic and Chaotic Behaviour*. Cambridge University Press, New York, 2004.
9. M. Maeda, S. Lu, G. Shaulsky, Y. Miyazaki, H. Kuwayama, Y. Tanaka, A. Kuspa, W. Loomis: Periodic signaling controlled by an oscillatory circuit that includes protein kinases ERK2 and PKA. *Science*, 304 (2004), 875–304.
10. V. Manca, L. Bianco, F. Fontana: Evolutions and oscillations of P systems: Theoretical considerations and applications to biochemical phenomena. In *Membrane Computing. International Wrkshop WMC2004* (G. Mauri, Gh. Păun, M.J. Pérez-Jiménez, G. Rozenberg, A. Salomaa, eds.), LNCS 3365, Springer, 2005, 63–84.
11. V. Manca, L. Bianco: Biological networks in metabolic P systems. Submitted, 2006.

12. R. Milner: *Is Computing an Experimental Science?* Laboratory for Computer Science, University of Edinburgh, Report Series ECS-LFCS-86-1, August 1986, Edinburgh, 1986.
13. Gh. Păun: Computing with membranes. *J. Comput. System Sci.*, 61, 1 (2000), 108–143.
14. Gh. Păun: *Membrane Computing. An Introduction*. Springer, 2002.
15. L.A. Segel, I.R. Cohen, eds.: *Design Principles for the Immune System and Other Distributed Autonomous Systems*. Oxford University Press, 2001.
16. E.O. Voit: *Computational Analysis of Biochemical Systems*. Cambridge University Press, 2000.
17. The P Systems Web Page: *http://psystems.disco.unimib.it*.