

RESEARCH QUESTIONS IN STATE TRANSITION MODELS OF BIOMOLECULAR DYNAMICS

G. FRANCO* and V. MANCA⁺

*Department of Computer Science, University of Verona,
Strada Le Grazie, 15, Verona, I-37134, Italy*

**E-mail: franco@sci.univr.it*

*⁺ vincenzo.manca@univr.it
www.di.univr.it*

G. SCOLLO

*Department of Mathematics and Computer Science, University of Catania,
Viale A. Doria, 6, Catania, I-95125, Italy*

E-mail: scollo@dmf.unict.it

www.dmi.unict.it

Recent developments of discrete models to analyse biological processes motivate the revisitation of typical concepts of classical dynamics in a completely discrete context. This work aims at identifying a few relevant problems outstanding in biomolecular computing, and at outlining research directions and strategies for their solution.

Keywords: State transition dynamics; Metabolic P systems; MP graphs.

1. Introduction

The classical approach to study dynamical systems is focused on differential equations, that impose local (infinitesimal) relations on quantity variations from which, under suitable hypotheses, one can analytically reconstruct the global dynamical behaviour of the system. Nevertheless, recent developments of discrete models to analyse biological processes motivate the revisitation of typical concepts of classical dynamics in a completely discrete context,³ such as that provided by cellular automata,²⁵ symbolic dynamics,^{11,12} or, more generally, by state transition systems.^{17,24}

Namely membrane systems (or P systems), introduced as a distributed computational model inspired by the structure and the functioning of the

living cell,^{18,19,22} are currently investigated as a promising theoretical model of biochemical and cellular processes.^{1,4,8,16,20} They were extensively investigated in terms of their computational power and efficiency, while a more recent perspective points out the nature of such systems as dynamical systems, thus moving the interest from halting configurations, to trajectories, attractors, periodic and fixed points. Dynamics of discrete systems, besides being instrumental to the analysis of metabolic processes in the context of P systems,^{13,15} in general prove most natural in the representation of biomolecular dynamics, where the symbolic entities which come into play are easily amenable to strings or related structures, ranging from multisets to dynamical networks.^{3,26}

This work aims at identifying a few relevant problems outstanding in biomolecular computing,⁶ and at outlining research directions and strategies for their solution. The next three sections, which give a contrived summary of basic concepts of state transition dynamics and metabolic systems, along with relevant research results, form the background for the subsequent outlook into the possible future of this research.

2. State transition dynamics

In Ref. 17, we addressed the problem of considering, in general terms, dynamical systems that are completely discrete in that, not only are the time instants natural or integer numbers, but also the space is a discrete entity. We characterized transitions and space states of a discrete dynamical system by means of *state transition dynamics*, defined as a pair $\langle S, q \rangle$, where S is a set of states and q is a binary relation on S , the *transition relation* on states.

A state transition dynamics is *eternal* if the transition relation q is total. We assume this to be always the case in the systems of our interest, that is, the set of *final* states $S \setminus \text{dom}(q)$ is empty. Every dynamics may be easily extended to an eternal one by turning each final state into a fixed point of the transition relation, that is by adding a transition from that state to itself in the extended dynamics.

As in Ref. 17, we call *quasistate* any subset of S . With the relation algebra notation from Ref. 24, the *successor* quasistate of state x under q is written $\text{img}(x; q)$.

This is generalized to denote the successor of quasistate x under q , defined as the union of the successor quasistates of states in x . The *orbit* of origin x , or x -orbit, results from iteration of this concept, viz. it is defined as the infinite sequence of quasistates $(x_i \mid i \in \mathbb{N})$ such that $x_0 = x$ (or

$x_0 = \{x\}$, if x is an individual state) and $x_{i+1} = \text{img}(x_i; q)$. The orbit is *periodic* if there exists a positive number n such that $x_n = x_0$, while it is *eventually periodic* if, for some non-negative k , it evolves into a periodic one after a k -step *transient*, that is, if there exists a positive n such that $x_{k+n} = x_k$. A *basin* B is a nonempty quasistate that is closed under q , that is, $\text{img}(B; q) \subseteq B$. Every orbit gives rise to a basin, by taking the union of the quasistates which it consists of.

For any state x in S , a *trajectory* of origin x , or x -trajectory, is a function $\xi : \mathbb{N} \rightarrow S$ such that, with subscript argument, $\xi_0 = x$ and $\xi_{n+1} \in \text{img}(\xi_n; q)$. $\xi_{\mathbb{N}}$ denotes the image of this function. An x -flight is an injective x -trajectory. Any x -trajectory thus runs “inside” the x -orbit. The difference between these two concepts is apparent, but it practically disappears for *deterministic* dynamics, viz. those where the transition relation is actually a function—hence orbits are singleton sequences, and one may say *the* x -trajectory, for any state x .

In Refs. 17 and 24 the general case of nondeterministic dynamics is considered, whereby classical dynamics concepts such as *attracting set*, *attractor*, and *recurrence* come in two distinct modal flavours. While we refer the interested reader to the cited work for details about the formulations of these concepts, and related results, in the general, nondeterministic case, in the rest of this paper we shall confine ourselves to the deterministic case, where modal differences disappear and the aforementioned concepts take the following shapes.

A state x is *recurrent* if it occurs at least twice (hence infinitely often) in the x -trajectory. An *attracting set* A of a basin B is a nonempty subset of B such that, for every state x in B , the x -trajectory eventually gets into it; more precisely, $\forall x \in B \exists k \in \mathbb{N} : \forall n \geq k \text{ img}(x; q^n) \subseteq A$, where q^n denotes the n -fold iterated composition of q . An attracting set of B that is minimal under set inclusion is called the *attractor* of B . We may say *the* attractor since this is unique, for a given basin B , whenever it exists.¹⁷ The image of an x -flight is a simple example of a basin that has no attractor. The deterministic-case corollary of a result in Ref. 24 states that the attractor of a basin B exists iff B has no flights, and that in this case the attractor of B consists of the set of recurrent states in B .

In order to investigate the discrete dynamics of systems modeling biological phenomena, in next section we present the metabolic P systems, which are a novel alternative to classical P systems, having in particular a deterministic evolution.

3. Dynamics of P systems

P systems are based on the *compartmentalization of the workspace* and on *multiset rewriting*. These basic concepts have clear biological counterparts in the role that membranes play in biological organisms, and in the biochemical basis of any biological entity, respectively.

In the standard model, multisets of objects evolve under maximal parallel application of evolution rules in each compartment (called region) of the membrane structure representing the system. That means that a maximal number of applicable rules is non deterministically chosen to perform the transition of the system from one configuration to the next one. The rules act synchronously in the compartments and movement of objects is allowed, usually among adjacent regions. We refer to Refs. 18, 19 and 22 for details on the definitions of P systems structure and its most important variants.

A non-deterministic P system, with rules working in the maximal parallel way but with a priority given by the context, has been proposed in Ref. 8 as a model for the (cellular) healing process occurring in response to an internal injury at the knee tissue. The non-determinism of the model, which ensures in cases of serious injuries, has been the key point to describe the biological reality experimentally observed, where the healing process has different and non predictable outcomes, in some cases the injury is repaired while in other cases (even chronic) arthritis is primed.

However, from a biochemical viewpoint, the evolution strategies considered in standard P systems do not seem to be fully adequate. Indeed, in these systems where state transitions are determined by a maximally parallel application of multiset rewriting rules, the dynamics of a population of chemicals governed by biochemistry laws is not expressed, especially in its *regulatory* aspects, which involve a dynamical *change of strategy*, depending on the system state.

Namely mitotic oscillations are a mechanism exploited by nature to regulate the mitosis process, that is, the cell division aimed at producing two daughter cells identical to the single parent cell. Mitotic oscillations concern the fluctuation of activation state of the substances involved in the process. In these cases, the dynamics regulating the cyclic oscillation of some biochemicals is fundamental to figure out the regulation mechanisms that provide the life of the cellular system.

Therefore, alternative strategies of rule application for P systems were suggested (e.g., a probabilistic approach may be found in Ref. 21) in order to investigate and determine the dynamics of P systems (i.e., their evolu-

tion in time) as models of biochemical processes, especially with oscillating components.^{4,16} A novel approach is introduced by the model of *Metabolic P systems*,^{1,2,13,15} that addresses the question by equipping each multiset rewriting rule, also called *metabolic rule*, with a *reaction map*, a real-valued function of the system state that yields the “competitive strength” of that rule at the given state. Every reactivity (the value of the reaction map in a certain state of the system) denotes the ability of the corresponding rule to compete against other rules in capturing part of a population of objects, on which the reaction is performed.² The reactivity of a rule at a given instant depends on the state of the system, defined as the concentration and localization of all substances.

Metabolic rules thus compete for the allocation of biochemical resources, whose types occur in the rule pattern. Allocation of resources, which are available in finite amounts, is governed by a *mass partition principle*, where reaction maps determine the relative allocation factors. Each metabolic rule acts on a system state by consuming/producing amounts of biochemical resources that are integer multiples of a *reaction unit* for that rule in the given state, thus generalizing the notion of molar unit (Avogadro’s principle). In particular, according to its stoichiometric “reading”, any rule determines its own reaction unit and therefore the amount of substances which it consumes and produces.

An example may be useful to clarify these principles. Let $\Sigma = \{A, B, C, \dots\}$ be an alphabet of biological species (or types), we define $q : \Sigma \rightarrow \mathbb{N}$ 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, then a partition strategy for species A is necessary. If a, b, c are the concentrations of species A, B, C respectively, then the reactivities associated, in a state q , to rules r_2, r_3, r_5 and r_7 , which ask for A molecules, could be:

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

We define

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

as the *total pressure* on A . Then, for each of the competing rules r_j we consider the *partial pressure* (or *weight*) of r_j on type A as

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

In this example, 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 factors of the amount of species A , available in state q , relatively to those rules which need objects A for performing their reactions.

At the end, let us assume that one of the rules competing for A , say rule r_2 , has the following form: $AAB \rightarrow AC$, and let us suppose that, according to the point (3) expressed above, n objects of type A were allocated to r_2 and m objects of type B were allocated to r_2 too. The corresponding reaction unit turns out to be

$$u_{r_2} = \min\{n/2, m\}$$

and this means that $2u_{r_2}$ objects of type A and u_{r_2} objects of type B are consumed, while u_{r_2} objects of type A and C are produced.

This globally states that u_{r_2} objects of type A and u_{r_2} objects of type B are replaced by u_{r_2} objects of type C . But, the important thing to point out here is that rule r_2 is absolutely different from a rule r' having the form $AB \rightarrow C$, and this is due to the fact that the two rules imply different competition factors, and consequently, different mass partitions. As a matter of fact, in the second case the reaction unit would have been $u_{r'} = \min\{n, m\}$.

In general, if S_r is the set of substances which are reactants of rule r , and if we set:

$$R(Y) = \{r | Y \in S_r\}$$

and

$$K_{Y,q} = \sum_{r \in R(Y)} f_r(q)$$

then the reaction unit $u_r(q)$, of rule r in state q , is given by

$$u_r(q) = \min\left\{\frac{q(Y) \cdot f_j(q)}{K_{Y,q}} \mid Y \in S_r\right\}.$$

The strategy outlined here suggests a natural representation of rules as graphs with two levels. The first level describes the reaction itself (the *stoichiometry*, that is the network of physical connections between species),

while the second level expresses the layout of *regulation* which tune the relative strengths of rules. The formal representation of biological networks is the subject of the following section.

In the notationally simpler case of a system with no workspace partitioning, *i.e.* with only one (outer) membrane, states are functions of type $\Sigma \rightarrow \mathbb{N}$, where Σ is the finite alphabet of biochemical resource types, and a deterministic dynamics on such states is computed, for a given set R of metabolic rules on Σ equipped with reaction maps $F = (f_r \mid r \in R)$, by the *P metabolic algorithm* (PMA),¹³ which formalizes the resource allocation strategy outlined above. A generalization of the PMA to P systems with inner membranes seems feasible, as well as to such systems with possible resource flow through membranes, but at the cost of increasing notational complexity.

4. MP Graphs

While the investigation into state transition dynamics carried out in Refs. 17 and 24 is motivated by the wish to consider dynamical systems where not only time but also the state space is discrete, it is to be noted that actually no assumption about state space structure is made in the definitions and results obtained in that work. As a matter of fact, one may easily specialize state transition dynamics by making additional assumptions about state space structure, whether discrete or continuous as it proves convenient. Thus, for example, even the simple case of metabolic P systems without inner membranes may get a richer state space structure just by taking states to be real-valued functions on the type alphabet, if continuous values prove more convenient to express mass quantities. Furthermore, regardless of the discrete or continuous nature of the functions' codomain, the state space of metabolic P systems is easily endowed with a metric, *viz.* it is a finite-dimensional vector space on that codomain (since the type alphabet is finite).

A class of questions arises from a valuable feature underlying the very design of metabolic P systems. Here, two distinct levels of description of biochemical dynamics are readily recognized: the *stoichiometry* level, represented by the metabolic rules, and the *regulation* level, represented by the reaction maps assigned to them. Bidirectional interaction between the two levels is inherent, since state transitions are determined according to the given stoichiometry, but with relative strength determined by the regulation, while, conversely, regulation depends on the state.

The formalism of *MP graphs* was introduced in Ref. 15 in order to repre-

sent such two levels of a system. They are graphs with three types of nodes, corresponding to the substances, to the rules, and to the reaction maps, and two types of arcs, one connecting substance nodes and rule nodes, and the other one connecting substance nodes to reaction map nodes. An example of MP graph is displayed in Fig. 1, excerpted from Ref. 7 (but turned to black and white because of editorial constraints), where the interruption of the mitosis after a DNA damage, due to degradation of Cdc25A phosphatase, was modeled.

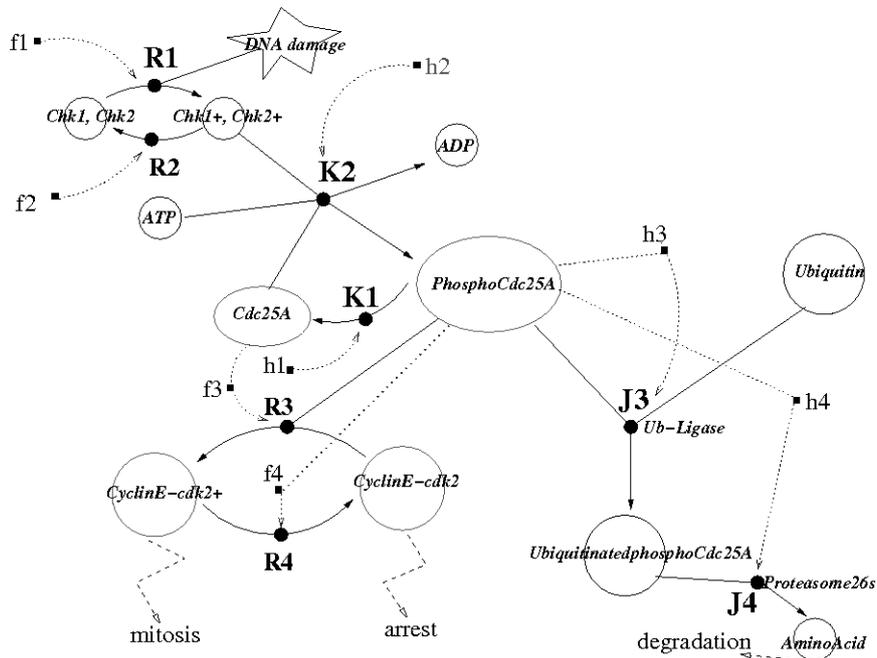


Fig. 1. Metabolic P graph of the process leading to a p53 independent G1 arrest.

The formalism of metabolic P graphs highlights the relevant information of the biological network dynamics, and the individuation of few parameters rules the basic mechanisms, for example of Cdc25A degradation, involving a couple of important mitotic oscillators.

A first, basic question is: For a given dynamical *behaviour* of a biochemical system, how can one find the metabolic rules and reaction maps that best reproduce that behaviour? This may be called the general *simulation problem* for metabolic P systems. A somewhat “easier” (but: how much

really easier?) subproblem of the general one results from assuming that a solution for the stoichiometry level is available somehow, but nothing is known about the reaction maps. This is thus the general *regulation problem* for metabolic P systems. Finally (for the time being), one may happen to know everything about the stoichiometry of the given behaviour, and *something* about plausible reaction maps (*e.g.* their polynomial form), but without knowing the values of specific parameters (*e.g.* polynomial coefficients) which determine those maps uniquely. This is the *tuning problem*, a subproblem of the regulation problem as it were.

Different strategies for the solution of the simulation problem are exemplified in Ref. 15, where the mitotic oscillator model proposed by Goldbeter^{9,10} is taken as a case study. This model proves adequate to account for the simplest form of mitotic oscillations, as found in early amphibian embryos. Goldbeter's model consists of an autonomous system of ordinary differential equations. Three different metabolic P system models are developed in Ref. 15 for the case under study. Two of them are directly derived from Goldbeter's model under different "readings" of the differential equations in terms of metabolic rules and reactions maps; this clearly tells that, in general, the "translation" of a differential model into a metabolic P model is not unique. Furthermore, the third P model is directly synthesized from the biochemical description of the phenomenon under study; it exhibits the desired oscillatory behaviour as well as the two other P models do, and it is even a simpler model.

Moreover, a general relationship between MP graphs and Ordinary differential equations (ODE) was found in Ref. 5, where the transformation of an MP graph in an ODE systems is proposed in terms of mass action principle, and it is shown that given an ODE system E there exists an MP graph having E as associated ODE system by the previous transformation.

Finally, about the tuning problem, there are two strategies under investigation to systematically compute the reactivities of MP metabolic graphs. One of them goes along with genetic algorithms, which find "good" values in order to obtain a desired behaviour, while the other one deals with the method of *MP Log-gain Regulation*,¹⁴ that is based upon the computational progress of the MP algorithm starting by actual biological data.

In this regard, it seems relevant to investigate an approach which proposes linear operators to describe the constant part of the system. Namely, in Ref. 23, biochemical networks were studied by means of linear systems, as reaction equations define a linear stoichiometric conversion of substrates into products. The flux distribution in the system was given by a suitable

bases of the null space of the sparse stoichiometric matrix S associated to the system along with chemical elements and ionic charge balances. Such a linear system was solved by methods of linear programming, and the method was used for genome-scale metabolic maps.

5. Future Research

The situation outlined above raises interesting questions relating to further analysis of state transition dynamics for state spaces endowed with topological structure, such as the following ones: 1) Definability of weaker notions of recurrence, where one would replace *exact* recurrent occurrence of the given state in its trajectory with *approximate* occurrence, viz. recurrence of seeing the trajectory get across a sufficiently (or arbitrarily) small neighbourhood of the origin. 2) Definability of weaker notions of attraction, where, similarly to the previous question, exact inclusion of trajectories or orbits in the attracting set is replaced by approximate inclusion. 3) Which (if any) of the possible answers to the previous questions deliver a straightforward generalization of the characterization results, linking recurrence and attractors, already obtained in Ref. 24 for structureless state transition dynamics?

We believe that the modeling with P systems can prove to be an exciting new way to investigate discrete dynamics, where mathematical properties of the system can be hopefully associated to medical conditions.

References

1. L. Bianco, F. Fontana, G. Franco and V. Manca, P Systems for Biological Dynamics, in: G. Ciobanu, Gh. Păun, M.J. Pérez-Jiménez (Eds.), *Applications of Membrane Computing*, Springer, Natural Computing Series (2006) 81–126.
2. L. Bianco, F. Fontana and V. Manca, P Systems with reaction maps, *International Journal of Foundations of Computer Science*, vol. 17:1, 2006, 27–48.
3. C. Bonanno and V. Manca, Discrete dynamics in biological models, in: Gh. Păun, C. Calude (Eds.), *Romanian Journal of Information Science and Technology*, vol. 1-2:5, 2002, 45–67.
4. F. Fontana, L. Bianco and V. Manca, P systems and the modeling of biochemical oscillations, in: R. Freund, Gh. Păun, G. Rozenberg, A. Salomaa (Eds.), *Membrane Computing, International Workshop, WMC6, Vienna, Austria, 2005, Selected and Invited Papers*, Springer, LNCS 3850, (2006) 200–209.
5. F. Fontana and V. Manca, Discrete solutions of differential equations by metabolic P systems, *Theoretical Computer Science*, to appear.
6. G. Franco, *Biomolecular Computing – Combinatorial Algorithms and Laboratory Experiments*, Doctoral dissertation, University of Verona, Italy (2006).
7. G. Franco, P. H. Guzzi, V. Manca, T. Mazza, Mitotic Oscillators as MP

- Graphs, H.J. Hoogeboom et al. (Eds.) *Proceedings of the 7th Workshop on Membrane Computing*, Springer-Verlag, LNCS 4361, (2006) 382-394.
8. G. Franco, N. Jonoska, B. Osborn, A. Plaas, Modeling Knee Injuries by Membrane Systems, submitted, October 2006.
 9. A. Goldbeter, A Minimal Cascade Model for the Mitotic Oscillator Involving Cyclin and cdc2 Kinase, *PNAS*, vol. 88:20, 1991, 9107–9111.
 10. A. Goldbeter, *Biochemical Oscillations and Cellular Rhythms. The molecular bases of periodic and chaotic behaviour*. Cambridge University Press, New York, USA, 2004.
 11. P. Karp Pathway Databases: A Case Study in Computational Symbolic Theories, *Science*, vol. 293, 2001, 2040-2044.
 12. D. Lind, B. Marcus *An Introduction to Symbolic Dynamics and Coding*, Cambridge University Press, 2000.
 13. V. Manca, Topics and Problems in Metabolic P Systems, in: Gh. Păun, M.J. Pérez-Jiménez (Eds.), *Proc. of the Fourth Brainstorming on Membrane Computing*, Sevilla, Spain (2006), Fenix Editora.
 14. V. Manca, The Metabolic Algorithm for P Systems, Principles and Applications *Theoretical Computer Science*, to appear.
 15. V. Manca and L. Bianco, Biological Networks in Metabolic P Systems, *BioSystems*, to appear.
 16. V. Manca, L. Bianco and F. Fontana, Evolutions and oscillations of P systems: Theoretical considerations and applications to biochemical phenomena, in: G. Mauri, Gh. Păun, M.J. Pérez-Jiménez, G. Rozenberg, A. Salomaa (Eds.), *Membrane Computing*, Springer, LNCS 3365, (2005) 63–84.
 17. V. Manca, G. Franco, G. Scollo, State transition dynamics: basic concepts and molecular computing perspectives, in: M. Gheorghe (Ed.), *Molecular Computational Models: Unconventional Approaches*, Idea Group, Hershey, PA, USA (2005) 32–55.
 18. Gh. Păun, Computing with membranes, *J. Comput. System Sci.*, vol. 61:1, 2000, 108–143.
 19. Gh. Păun, *Membrane Computing. An Introduction*, Springer, 2002.
 20. M. Pérez-Jiménez, P Systems-Based Modeling of Cellular Signalling Pathways. *Proceedings of the 7th Workshop on Membrane Computing*, Springer-Verlag, LNCS 4361, (2006) 54-73.
 21. D. Pescini, D. Besozzi, G. Mauri, C. Zandron, Dynamical probabilistic P systems, *International Journal of Foundations of Computer Science*, vol. 17(1), 2006, 183-204.
 22. P Systems Web Page, <http://psystems.disco.unimib.it> .
 23. C. Schilling and B.O. Palsson The Underlying Pathway Structure of Biochemical Reaction Networks, *PNAS* vol. 95, 1998, 4193–4198.
 24. G. Scollo, G. Franco, V. Manca, A Relational View of Recurrence and Attractors in State Transition Dynamics, in: R.A. Schmidt (Ed.), *Relations and Kleene Algebra in Computer Science*, Springer, LNCS 4136, 2006, 358–372.
 25. S. Wolfram, *Theory and Application of Cellular Automata*, (1986) Addison-Wesley.
 26. A. Wuensche, Basins of Attraction in Network Dynamics: A Conceptual

Franco G. et al

12

Framework for Biomolecular Networks, in: G. Schlosser and G.P. Wagner (Eds.), *Modularity in Development and Evolution*, (2002), Chicago University Press.