A biochemistry-inspired artificial chemistry: LAC

José A. Pereira

***Abstract*—A computational artificial chemistry (AC) system, called LAC, was developed bearing in mind the major biochemistry principles. Basic elements may be interpreted as small carbon molecules capable of bonding by nucleophilic substitution or nucleophilic addition. Reactions similar to hydrolysis and oxidation-reduction complete the present set of LAC site interaction rules. Reaction extension is governed by artificial thermodynamic principles and the reaction rate is controlled by specific catalysts made of the same kind of basic elements they act upon. The preliminary results of LAC experiments show the ability to mimic enzyme regulation and oscillatory reactions.**

***Index Terms*—Artificial chemistry, catalysis, Lotka-Volterra mechanism, oscillating reactions.**

I. Introduction

Last century's breakthrough on biochemistry research, leading to an in-depth knowledge of cell's molecular organization, coincided with the development of new epistemological and formal approaches to explain the phenomenon of life, from Erwin Schrödinger's 1944 text "What is life" to the self-reproducing automata of John von Newmann [1], the Universal Turing Machines [2], Hofstadter's Typogenetics [3] and the autopoiesis theory of Maturana, Varela and Uribe [4]. The autopoiesis theory was perhaps the first wide-range attempt to develop formal criteria to generally define living systems. In this theory life is a system property, characterized by an organized component production network that also includes a physical boundary that individualizes it. This organization is operationally closed, depending on a number of specific internal relational constraints and it has the special ability of reproducing itself. Autopoiesis is a convenient guiding framework to use in the development of computational artificial life and it is applicable to an artificial chemistry setup [4, 5].

A variety of computational AC approaches have been used to investigate the fundamental aspects of living systems, reviewed in [6] and more recently analysed for some cases in [7]. To design an AC system it is necessary to decide on the nature of its three general components: the basic elements set, the set of rules that define interactions between elements and an algorithm for the time evolution of the whole system in a space. In some cases the criteria for the design of the AC components are purely abstract but in other cases the essential features of a preexisting system inspire the creation of the AC model components so that some isomorphism can be established, giving a more precise meaning to, and a straightforward interpretation of, the results [3]. As we prefer the latter approach, our aim was to develop an artificial chemistry closely inspired in the basic principles of chemistry and of biochemistry, that we called LAC, making as many parallels as possible to biological molecules functions and the general cell metabolic strategy. Although the autopoiesis theory has a formulation independent of thermodynamical criteria the question of energy was recognized as essential for the development of LAC and was addressed quantitatively.

Manuscript received August 29, 2005.

J. A. Pereira is with the Universidade do Porto at ICBAS - Instituto de Ciências Biomédicas de Abel Salazar, Largo Prof. Abel Salazar, 2, 4099-003 Porto, Portugal (phone: +351 222 062 288; fax: +351 222 062 232; e-mail: jpereira@icbas.up.pt).

II. Specifications for LAC

A. Basic elements

The basic elements should not represent anything more complex than a macromolecule although the small molecule level or the atom level would be better because it provides a finer description. Also, in biochemistry the shape of molecules is a most important feature because the whole cellular organization depends on specific interactions selected by molecular recognition. The aggregates of the basic elements, or compounds, should therefore be represented by some kind of data structure bearing interconnecting flags and internal coordinates for each element, like a simplified molecular graph [8], and should be recognized by its shape using pattern matching. This can be a source of novelty, in the sense defined in [9], if new shapes are interpreted by an AC system algorithm as corresponding to new functions. Such genotype-phenotype relationship insures closure between (bottom-up) compound construction and (top-down) chemical activity, in line with the concepts of self-reproducing automata or Turing machines.

B. Reaction rules

Still with biochemistry in mind, the LAC reaction rules have to account for at least two types of general interactions:

intermolecular association and covalent bonding. The former is not a real chemical reaction but its role in biochemistry is too important not to acknowledge in artificial chemistry because it is in the origin of molecular recognition. Covalent bonding in biological carbon chemistry can be tentatively restricted to three main types of reactions: nucleophilic substitution, nucleophilic addition and oxidation-reduction. The elements should therefore have a certain number of sites with Lewis acid-base properties that can react by simulated nucleophile-electrophile bonding and be modified by simulated redox reactions. A probability for each reaction will have to be calculated based on energy criteria.

Reactions in real chemistry are governed by statistical factors dependent on energy terms but the autopoiesis theory, for example, is consciously not quantitative (in fact, for its purpose, it doesn't have to be) and has to be complemented by an energy definition when designing a concrete system [10]. Any practical realization of an AC system forcefully implies the definition of reaction constraints that can always be interpreted as energy dependent. Consider that, in spite of the diversity of AC approaches in literature, there is a basic feature common to all: the emergence of some sustained organization pattern in the chemical reaction space resulting from the free evolution of the system, which was initially set in a more random state. It is known that, in the real world, a necessary thermodynamic condition to achieve such a state is that the system be kept far from equilibrium by continuously feeding it with high chemical potential elements (elements that tend to react) [11]. Organization is a manifestation of low entropy (high-energy) in a dynamical chemical system and can only be attained, and maintained, by fuelling it continuously, configuring an open and dissipative system. This principle holds for the above-mentioned AC systems that, although immaterial, show a behavior compatible with what is observed in the physical world. In LAC we opted for a quantitative definition of the artificial energy value associated with each type of reaction. The energy values were used to calculate the reaction probabilities by an exponential relationship similar to the Arrhenius equation for the rate constant of a reaction.

Catalysis is at the very heart of biochemistry and enzymes mediate virtually all cell metabolism. So, in LAC only catalyzed reactions are possible because controlled and specific catalysis is the biological way to generate kinetic constraints that contribute greatly to organization. The simpler model for catalysis is the Michaelis-Menten enzyme kinetics model that includes the formation of a complex between catalyst and the reactant compound (enzyme and substrate) previous to the reaction itself. Inhibition and activation of enzymes is also a critical aspect of biochemistry that LAC has to cover. So, each compound recognized by the system as a catalyst must have one catalytic site and from none to two regulatory sites attributed. Association with all these sites is selective and this selectivity is a function of the composition of the catalytic compound, as is its kinetic properties. The LAC system can qualitatively evolve by the appearance of new catalysts that specify new reaction rules.

C. *Dynamics and space*

A 2D or 3D space is convenient to study the emergence of physical boundaries by intermolecular association [12], but it is not a requirement to study the onset of chemical organization. The biological space is in fact three-dimensional but if the goal is restricted to represent concentrations then a well-stirred reactor (WSR) space is adequate. Dynamics is simple: one (for unimolecular reactions) or two elements are randomly drawn from the element pool and the reaction probability is evaluated. The reaction will occur only if a random number between 0 and 1 is equal or less than that probability. In an exclusive catalysis artificial chemistry like LAC the only potentially effective collisions should be between catalysts and other molecules recognized by it. All other collisions count for the time record but are elastic.

III. THE DESIGN OPTIONS FOR LAC

A. *Basic elements*

Basic elements are explicit data structures that have from none to four covalent bonding sites defining 90 or 180-degree angles in the same plane (orthogonal configuration). Compounds are also explicit but constructed by cross-referencing of identification numbers between elements. These also contain 2D internal coordinates coherent with its position on a compound. The bonding sites are of four different kinds, inspired in the organic functions carboxylic or phosphoric acids (kind 1), aldehyde (kind 2), hydroxyl or amine (kind 3) and methyl (kind 4). An element can be interpreted as a small molecule with any combination of these kinds of functions, defining an element type. As in the Lewis acid-base concept, kind 1 acts as an acid or a base, kind 2 as an acid, kind 3 as a base and kind 4 doesn't have any Lewis acid-base properties.

An analogy between the elements in LAC and real chemical elements cannot be made because LAC elements can sometimes be chemically transformed.

B. *Reaction types between bonding sites*

All reactions conserve the number of basic elements and sites. Nucleophilic addition reactions involve only kind 2 sites as acid (a) and kinds 1 or 3 as base (b) and are of the form:

$$a + b \rightleftharpoons ab \tag{1}$$

This is analogous to the formation of a hemiacetal or a Shiff base in carbon chemistry.

Nucleophilic substitution (SN) is a substitution of a base (b) by another base (v) in an acid (a). The acid may only be of kind 1 and the base may only be of kinds 1 or 3. The SN-type reactions are of the general form:

$$ab + v \rightleftharpoons av + b \tag{2}$$

Hydrolysis is the breaking of a bond between an acid (a) of kind 1 and a base (b) of kind 1 or 3. The reaction form is:

$$ab \rightleftharpoons a + b \quad (3)$$

Reaction (3) is related to a SN reaction in the sense that it configures a nucleophilic substitution by water. Water is therefore implicit in LAC.

Oxidation-reduction (redox) reactions do not form or break bonds but it transforms the kind of two bonding sites simultaneously by the general reaction:

$$x + w \rightleftharpoons y + z \quad (4)$$

in which the kind x is transformed in to kind y and, simultaneously, kind w is transformed in to kind z. The pairs x/y and w/z are akin to redox pairs. The reduced or oxidized sites may be involved in covalent bonds or not. A sequence was established for the oxidation state of a kind of bonding site, kind 1 being more oxidized and kind 4 less oxidized (more reduced). In a redox reaction each site may only be transformed from its kind into a neighbor kind in this sequence.

A fifth reaction type was defined, the rearrangement reaction, which in fact is not independent from the redox reaction just described but it has an equivalent in biochemistry and simplifies some chemical paths. Rearrangements are simply the switching between the kinds of two of a basic element sites.

C. Catalysis

A catalyst is just a LAC compound that meets certain structural specifications, like certain composition (number and type of basic elements), certain sequences or certain configurations. The functional properties of the catalysts are implicit, attributed by a high-level manager, and can be predefined or randomly related to certain compounds as they appear. The specific properties of a catalyst are: the type of reaction that it catalyses, the substrate and modulator compounds it recognizes and its kinetical properties. Catalysts in LAC specify rules of interaction because they add compound recognition (specificity) to the basic site reactions described in Section III-B. For the general mechanism of catalysis the Michaelis-Menten (M-M) model was adopted:

$$E + S \rightleftharpoons ES \longrightarrow E + P \quad (5)$$

where E is the catalyst, S is the substrate and P is the product. Some reactions defined in Section III-B are bimolecular but still fit the M-M model because only one of the reactants explicitly associates with the catalyst. Fig. 1 summarizes the energy profile for any reaction. The energy peeks correspond to reaction transition states that are not actually present as compounds in the simulations. The transition energies, denoted with the superscripted "‡", serve only to define different reaction probabilities. The energy ΔG_d is the only energy difference that is not dependent on the catalyst but solely on the reactants and products while all other energy differences are part of the specific catalyst description. The probability of transition between levels is calculated by the exponential expression $p = e^{-\Delta G/T}$, where ΔG is the positive energy difference between some initial level (reactants, products or intermediates) and the transition state that stands in the way to an adjacent energy level. The quantity T can be interpreted as the temperature of the system and serves as a general moderation constant that defines the steepness of the relationship between the probability and ΔG.

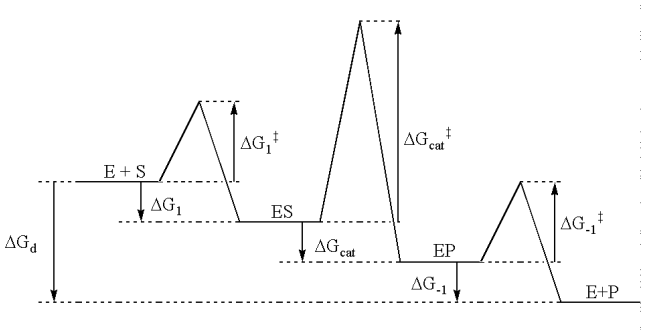

Fig. 1. General energy model used in LAC for catalyst action. The EP complex is not in (5) but stands for ES when the reaction goes in the reverse direction (from right to left). The top edges refer to transition states that do not actually exist in the ensemble of LAC compounds.

D. Energy

Energy differences for each reaction type and each catalyst are explicitly tabulated in a data structure of LAC for the experiments described in the Results section. However, energy differences could be calculated by using some algorithm dependent on reaction site combination on the colliding elements and the structure of reacting compounds.

The physics of LAC is characterized by a mixture of continuous and discrete constraints because reaction rates depend on the probability of random encounters of discrete compounds or elements (dependent only on concentrations) but also on a continuous probability function, $p = e^{-\Delta G/T}$ (since any energy difference ΔG is allowed) that depends on the energy of a virtual transition state.

IV. Results

Here are presented the results of three LAC experiments that demonstrate the convenience of the approach to mimic enzyme behavior in biological chemistry. In the following, a letter represents a basic element type; adjacent letters mean that the elements are strongly bonded. Compounds separated by a slash identify complexes, formed by weak bonding between catalyst (left of the slash) and substrate (right of the slash). In all experiments time is defined as the total number of element random choices divided by the total number of elements.

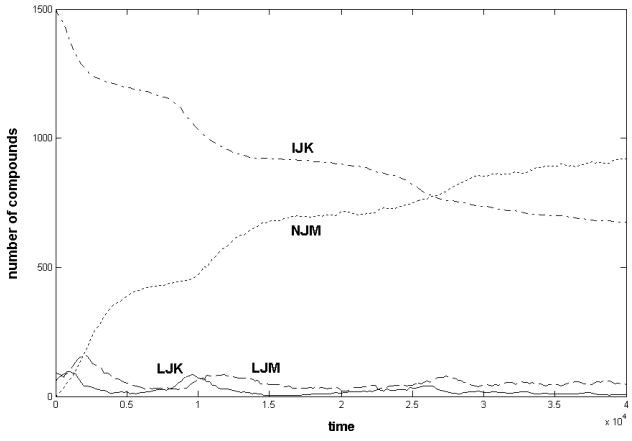

Fig. 2. Time evolution of reactions (6) to (8) in a closed system. Complexes are not shown. Damped oscillation of intermediate compounds quantities is observed.

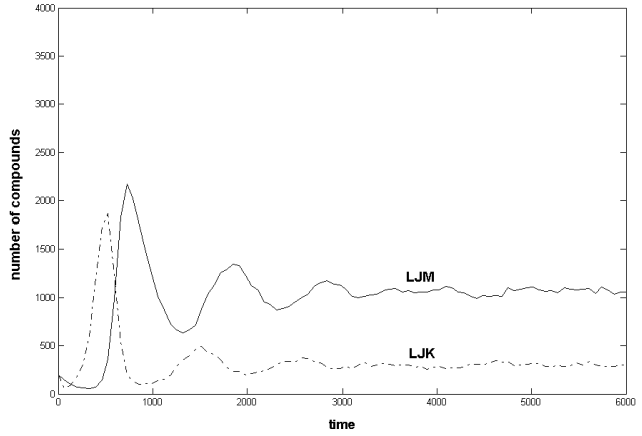

Fig. 3. Time evolution of reactions (6) to (8) intermediate compounds LJM and LJK in an open system. Exchange rate is 200 cycle^{-1}. Reactant IJK, product NJM and complexes are not shown.

Therefore, the time unit, or cycle, corresponds to a number of element choices equal to the total number of elements. This number was always kept constant throughout each experiment, even when the element pool was open to outside exchanges.

A. Implementation of the Lotka-Volterra model with catalysts also as substrates in an irreversible set of reactions

The classical Lotka-Volterra (L-V) autocatalytic model is known to be applicable to a wide range of phenomena, from chemistry and biology to economics [13]. The L-V chemical reaction system shows typical intermediate concentration oscillations for certain initial conditions and can be implemented in LAC as follows, with some compounds acting both as catalysts and substrates:

$$\text{IJK} + \text{LJK} \longrightarrow 2\ \text{LJK} \qquad (6)$$

$$\text{LJK} + \text{LJM} \longrightarrow 2\ \text{LJM} \qquad (7)$$

$$\text{LJM} + \text{LJN} \longrightarrow \text{NJM} + \text{LJN} \qquad (8)$$

Only compounds starting with the L element are catalysts, in this experiment. The sum of reactions (6) to (8) is simply

$$\text{IJK} \longrightarrow \text{NJM} \qquad (9)$$

The reactions described in (6) and (7) are autocatalytic because its own catalysts are reaction products, LJK for (6) and LJM for (7). The starting reactant is IJK and the final product is NJM. Reactions are all set to be virtually complete (irreversible), with large negative values for the energy differences ΔG_d. The reaction type used in all three reactions of this example is the rearrangement reaction that changes element types by swapping the position of different bonding sites in an element. When catalyst LJN is set much faster than LJM and LJM much faster than LJK an oscillatory pattern for LJM and LJK quantities is observed.

Fig. 2 shows the time evolution of a closed L-V system towards equilibrium. Transient oscillations are observed until the IJK quantity is too low and the potentially remaining oscillations are too small to stand out of the stochastic noise. The numbers of intermediate compounds LJK and LJM oscillate in time as the system evolves to equilibrium. Due to this oscillation the numbers of reactant IJK and product NJM change in a stepwise manner. An experiment with the same chemical system but with a constant influx of IJK and equal efflux of NJM (exchange rate of 200 cycle^{-1}), configuring an open system, showed the evolution represented in Fig. 3. The oscillations of intermediate compounds are now faster but still damped and tending to a non-equilibrium steady-state. The equilibrium quantity of LJM and LJK should be near zero, because the reactions are set to be virtually complete, but, in this open system experiment, the number of these compounds stabilize far from that value keeping the system far form equilibrium.

B. Allosteric catalyst activation by the product of a reversible SN reaction

Allosteric regulation of enzymes is a very important aspect of biochemical systems organization. To show the allosteric regulation of a LAC catalyst it was used a SN type reaction characterized by a ΔG_d of zero (reversible). The reaction

$$\text{D} + \text{pp} \rightleftharpoons \text{Dp} + \text{p} \qquad (10)$$

was catalyzed by a three element compound, LKK, in a much lower initial quantity than the reactants D or pp. The reaction steps, with intermediate reactions involving the catalyst, is

$$\text{LKK+ pp} \rightleftharpoons \text{LKK/pp} \qquad (11)$$

$$\text{LKK/pp} + \text{D} \rightleftharpoons \text{LKK/Dp} + \text{p} \qquad (12)$$

$$\text{LKK/Dp} \rightleftharpoons \text{LKK+ Dp} \qquad (13)$$

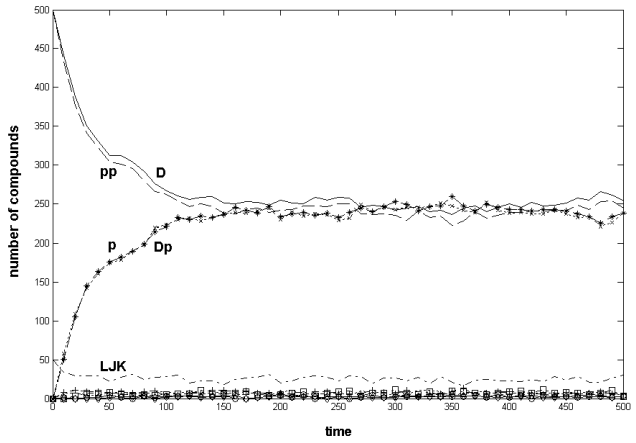

Fig. 4. Time evolution of reaction (10) to equilibrium. A hyperbolic shape, typical of a non-regulated catalysis is apparent. The catalyst LJK is not saturated. The intermediate reaction complexes are in very small number, at the bottom of the plot.

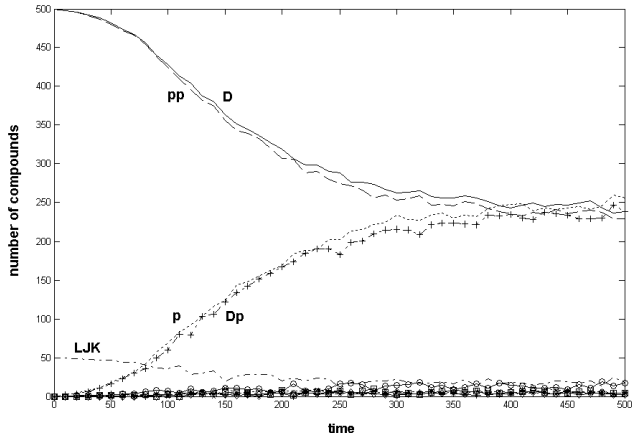

Fig. 5. Time evolution of reaction (10) but now the catalyst LJK is activated by product p. The result is a sigmoidal curve, dependent on the quantity of element p. The free LJK is in smaller numbers than in Fig. 4 due to association with p in the regulatory site.

The sum of (11), (12) and (13) give (10). In the experiment represented in Fig. 4 the catalyst is not regulated and so the reactant and product curves are of hyperbolic form. Equilibrium is attained with equal quantities of reactants and products, as expected for an ΔG_d of zero. A certain amount of the free catalyst and its complexes with the substrates pp and Dp are also visible. The time evolution of the same system but with a new kinetical definition for the catalyst LJK is shown in Fig. 5. The catalyst has now a very low activity unless the product element p associates in a regulatory site. The shape of the quantity curves are now sigmoidal, in agreement with what is typical for this regulatory condition in biochemistry, sometimes associated with cooperativity between enzyme subunits. The equilibrium stationary state is exactly the same as without regulation, as was expected. The reason for this is that only the kinetical properties of the catalyst were changed, not the reaction energy ΔG_d.

C. Sustained oscillations: evolution towards a limit cycle by allosteric activation of a LAC catalyst

Phosphofructokinase (PFK) is an allostericaly regulated enzyme of glycolysis and is the source of metabolic sustained oscillations [14, 15]. The PFK catalyzed reaction is an ATP phosphorilation of fructose-6-phosphate (F6P) that yields fructose-1,6-biphosphate and ADP. Phosphofructokinase is allostericaly activated by ADP, signaling a low cell's energy level.

Previous computer simulations of PFK reaction oscillations by conventional differential equations techniques were based on elaborations of the chemical model

$$\text{PFK} + \text{A} \longrightarrow \text{PFK} + \text{B} \qquad (14)$$

$$\text{LII} + \text{B} \longrightarrow \text{LII} + \text{H} \qquad (15)$$

where reactions are irreversible, PFK and LII are catalysts, and PFK is activated by B (the reactants A and B can be associated with F6P and ADP.) It is also assumed that product H continuously leaves the system with a constant rate ν and that A continuously takes its place by entering the system at the same constant rate.

The LAC simulations reported here use this model based on a rearrangement reaction type. After some search for adequate conditions, like initial quantities and enzyme properties, self-sustained oscillations of elements A and B quantities in near-anti-phase were observed, with amplitude and frequency dependent on the H to A exchange rate, ν. For each ν, however, some uncertainty was associated with frequency and amplitude of oscillations, as can be observed in Fig. 6. A larger population of elements diminishes this effect and a smaller population increases it. Nevertheless, a Fourier transform of A and B data showed that, for repeated runs with the same ν there is always the same predominant frequency of oscillation, defining therefore a limit cycle of oscillations, although noisy. A typical run is presented in Fig. 6 with ν equal to 10 cycle^{-1}.

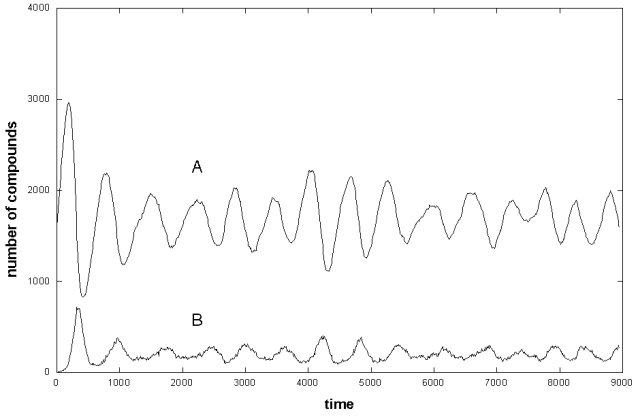

Fig. 6. Oscillations of elements A and B quantities for the reaction system (14)/(15) when ν is 10 cycle^{-1}. Initial quantities where: A=1600, B=0, H=12000, PFK=800 and LII =600.

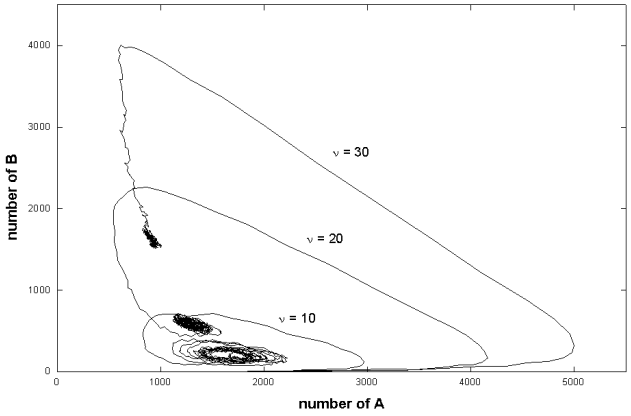

Fig. 7. Phase plane representation of A and B quantities evolution in reaction system (14)/(15) for exchange rate ν equal to 10 (the same experiment as in Fig. 6), 20 and 30. Initial quantites are all the same as reported in Fig. 6 caption.

Increasing or decreasing ν from 10 brings the system from an unstable (oscillating) steady-state to a stable non-equilibrium steady-state. In Fig.7 are represented the trajectories of A and B quantities in the phase plane where the limit cycle obtained for ν = 10 gives place to a stable (but also noisy) steady-state when ν is increased to 30.

V. Conclusions

As the results reported above show, the time evolution of reactions in LAC display the characteristics normally observed in real chemistry and differential equations simulations. The LAC system can be easily setup to mimic enzyme activity and shows a predictable, real chemistry, behavior, based on a quantitative definition of (artificial) energy and transition state models. On the other hand it is still an artificial chemistry system and should be able to model sustainable metabolisms and template compound replication, leading to the evolution of organized informational chemical systems. This line of work is now being pursued.

The obligatory catalyst paradigm in LAC is not a severe limitation. It was possible, for example, to implement the 2D Lotka-Volterra model without any adaptations to the program code. Chemical organization is hardly conceivable without specific regulated catalysts, and theories of prebiotic chemical evolution leading to life have to postulate the presence, at some point, of some catalytical substances, be it metal ions, minerals or some kind of carbon based polymers.

Our first goal was to design an AC system with a clear and as fine as possible formal likeness to chemistry and biochemistry that permitted a more straightforward explanation, if not a simulation, of the phenomenon of life as we know it, in line with the views expressed by Rasmussen et al. [16, pag. 346]. We believe that, as implemented in LAC, a clearer thermodynamic infrastructure for reaction rules and a more detailed basic elements and reactions definitions will greatly contribute to that effect.

Acknowledgment

To Pedro Russo and Ricardo Gama for the kick-off enthusiasm and for introducing me to Matlab.

References

[1] J. von Neumann, *Theory of Self-Reproducing Automata* (edited and completed by A. W. Burks), Urbana: University of Illinois Press, 1966.

[2] T. Ikegami and T. Hashimoto, "Active Mutation in Self-Reproducing Networks of Machines and Tapes," *Artificial Life*, 1995, vol. 2, pp. 305–318.

[3] D. R. Hofstadter, *Gödel, Escher, Bach: An Eternal Golden Braid*. New York: Basic Books, Inc., 1979, ch. 2 and 6.

[4] F. J. Varela, H. Maturana, and R. Uribe, "Autopoiesis: The organization of Living Systems, its Characterization and a Model," *Biosystems*, 1974, vol. 5, pp. 187–196.

[5] B. McMullin and F. J. Varela, "Rediscovering Computational Autopoiesis," in *Fourth European Conference on Artificial Life*, P. Husbands and I. Harvey, Ed. Cambridge, MA: MIT Press, 1997, pp. 38–47.

[6] P. Dittrich, J. Ziegler, and W. Banzhaf, "Artificial Chemistries - A Review," *Artificial Life*, 2001, vol. 7, pp. 225–275.

[7] H. Suzuki, N. Ono, and K. Yuta, "Several Necessary Conditions for the Evolution of Complex Forms of Life in an Artificial Environment," *Artificial Life*, 2003, vol. 9, pp. 153–174.

[8] G. Benkö, C. Flamm, and P. F. Stadler, "A Graph-Based Toy Model of Chemistry," *J. Chem. Inform. and Computer Science*, 2003, vol. 43, pp. 1085–1093.

[9] D. Groß and B. McMullin, "The Creation of Novelty in Artificial Chemistries," in *Artificial Life VIII*, Standish, Abbass and Bedau, Ed., 2002, pp. 400–409.

[10] K. Ruiz-Mirazo and A. Moreno, "Basic Autonomy as a Fundamental Step in the Synthesis of Life," *Artificial Life*, 2004, vol. 10, pp. 235–259.

[11] G. Nicolis and Y. Prigogine, *Self-Organization in Non-Equilibrium Systems*, New York: Wiley, 1977.

[12] B. Mayer and S. Rasmussen, "Dynamics and Simulation of Micellar Self-Reproduction," *Int. J. of Modern Physics*, 2000, vol. 11, pp. 809–826.

[13] L. Cairó, "Darboux first integral conditions and integrability of the 3D Lotka-Volterra system," *J. Nonlinear Math. Physics*, 2000, vol. 7, pp. 511–531.

[14] A. Goldbeter, *Biochemical oscillations and cellular rhythms*. Cambridge: The Cambridge University Press, 1996, ch. 2.

[15] R. Heinrich and S. Schuster, *The regulation of cellular systems*. New York: Chapman & Hall, 1996, ch. 2.

[16] S. Rasmussen, N. A. Baas, B. Mayer, M. Nilsson, and M. W. Olesen, "Ansatz for Dynamical Hierarchies," *Artificial Life*, 2001, vol. 7, pp. 329–353.