The role of energy in a stochastic model of the emergence of autocatalytic sets

Alessandro Filisetti[1], Alex Graudenzi[1], Roberto Serra[1,2], Marco Villani[1,2], Davide De Lucrezia[1], and Irene Poli[1,3]

[1]European Centre for Living Technology, S.Marco 2940, 30124, Venice (IT)
[2]Department Communication and Economics, University of Modena and Reggio Emilia, Via Allegri 9, 42100 Reggio Emilia (IT)
[3]Department of statistics, University Ca'Foscari of Venice, 30121 Venice (IT)
alessandro.filisetti@ecltech.org

Abstract

In most theories concerning the origin of life autocatalytic sets are supposed to play an important role in the phase transition between non-living and living matter. Although several theoretical models describe this phase transition, it is very hard to recreate the experimental conditions in wet lab. We here introduce a stochastic model of catalytic reaction networks with energy constraints, devoted to the study of the emergence of autocatalytic sets, in which some of the assumptions of the already existing model are relaxed in order to explore the possible reasons which make the emergence of autocatalytic cycles difficult or which make them unstable. Moreover, since living systems operate with a continuous exchange of matter and energy with the environment, we investigate the effects on the model behavior of changes in the rate of the energy intake.

Introduction

The life as we know today is the result of billions of years of evolution and, even though the first forms of life were simpler than today, a certain degree of complexity was surely necessary in order to lead off the phase transition between non-living and living matter.

Although different scenarios for the onset of life have been proposed[1], autocatalytic sets of molecules (ACSs) are considered of paramount importance to both extant biological systems and during the transition from non-living to living systems.
In the first case, ACSs represent the basic architecture of some of the most fundamental metabolic processes such as the citric acid cycle urea cycle, calvin cycle and beta-oxidation (Alberts et al., 2005), on the other hand, the emergence of ACSs might have played a pivotal role in acquiring autonomy and homeostasis during the emergence of the first living systems (Ruiz-Mirazo and Mavelli, 2008) and they have been regarded ever since as the blueprint of primeval living systems (Fishkis, 2007; Ma et al., 2010).

Though the RNA world scenario, with particular regard to the role of the ribozymes (Gilbert, 1986; Talini et al., 2009), provides a plausible solution with respect to the prebiotic storage and replication of the information, it has been proven that template-dependent polymerization can occur only for relatively short nucleotides strands catalyzed by remarkably long RNAs (Bartel and Unrau, 1999; Bartel, 1999).
On the other hand, looking at the metabolism-first approach, the self-replication of only a single catalyst is plausible only within a very complex chemical system.

In his theory concerning the emergence of ordered structures and patterns of activation from disordered interactions (the so called "order for free" hypothesis (Kauffman, 1993)) Stuart Kauffman pointed out the idea that all that is needed is the presence of a set of molecules composed of a sufficient number of different molecular types in which each molecule catalyzes a step in the formation of one or more other molecules in the set; then the catalytic closure is reached if each molecule in the set is catalyzed by at least one other molecule of the set. Based on a combinatorial approach, Kauffman stated that the emergence of autocatalytic sets is inevitable when the molecular diversity reaches a certain threshold (Kauffman, 1986).
While the Kauffman initial approach is based on an analysis of the reaction graph, without taking into account the dynamics, there are several models in the literature that study autocatalytic systems from a dynamical point of view, such as those by Dyson (Dyson, 1985), Eigen and Schuster (Eigen and Schuster, 1977), Kauffman (Kauffman, 1986), Farmer and Bagley (Bagley and Farmer, 1992; Bagley et al., 1989), Jain and Khrishna (Jain and Krishna, 1998), Lancet (Segre et al., 1998) and Kaneko (Kaneko,

[1]The main theoretical frameworks can be divided in the "gene first" approach, based on the template matching (Gilbert, 1986; Müller, 2006; De Lucrezia et al., 2007; Anastasi et al., 2007; Talini et al., 2009; Rios and Tor, 2009; Budin and Szostak, 2010), the "metabolism first" approach, based on the self-organization of the chemicals involved in (Oparin, 1924; Miller, 1953; Eigen and Schuster, 1977; Kauffman, 1986; Mossel and Steel, 2005; Lee et al., 1996; Saghatelian et al., 2001; Lifson, 1997) and the lipid-world (Segre et al., 1998)

2006).

Even if all these models predict the emergence of an autocatalytic set, observing it in a wet lab experiment remains a very difficult task. On the one hand, it is possible that the simplifications introduced by the in-silico models are unrealistic with respect to the extant biological systems but, on the other hand, the indications provided by the theory may be not correctly implemented in actual experiments.

In previous works (Fuchslin et al., 2010; Filisetti et al., 2011a,b, 2010) we introduced a novel stochastic model devoted to the study of the generic proprieties of catalytic reaction networks based on a particle description of the system, while in the present work we investigate the effects of the introduction of energetic constraints in the system.

Living systems cannot operate isolated from the environment and they need a continuous flow of energy and matter in order to be maintained far from the equilibrium. While the incoming flux of matter is necessary in order to feed the system with the elementary nutrients to be transformed in more complex molecules, energy is channeled into the construction of molecules whose constitutive reactions are energetically unfavorable. Energy is stored as chemical bonds in molecules called "carrier molecules", which diffuse rapidly in the cell and thereby carry energy from places of energy generation to the reactions requiring energy to occur (Alberts et al., 2005).
While results concerning the influence of different composition of the incoming flux of matter have been presented in (Filisetti et al., 2010; Fuchslin et al., 2010; Filisetti et al., 2011a,b), here we focus on the role of the energy, some first indications can be found in (Fuchslin et al., 2010), within a system composed of both energetically unfavorable and favorable reactions.

It is important to remark that, coherently with the scientific approach typical of complex systems biology (Kaneko, 2006), we are not interested in investigating the specific nature of the chemicals present in out model, nor the particular interactions among them, but rather in the characterization of the dynamical behaviors emerging from the interaction of a set of chemicals and in the detection of possible generic properties of this kind of systems.

In section II the principal features of the models are presented while in section III we describe how the energy has been introduced in the stochastic model. In section IV we discuss some preliminary results of a set of simulations in which we study the influence of the the amount of energy introduced in the system and, in the final section, conclusions and indications for further works are provided.

Description of the model

An exhaustive description of the model can be found in (Filisetti et al., 2011a) and (Filisetti et al., 2010); we here summarize the principal features for a better comprehension of the paper.
Taking inspiration from the original work by Kauffman (Kauffman, 1986) the principal entities of the model are linear chains, species from now on, oriented from left to right, composed of the concatenation of letters from a finite alphabet, e.g. [A,B] or [A,G, C, T].
Let X stands for the entire set of species and $x_i, i = 1, ..., N$, representing each single species. In accordance with the stochastic nature of the model the total amount (quantity of *molecules*) of each species x_i is denoted by $\hat{x}_i$. Since the reactions take place in a well-stirred tank reactor with fixed volume, the relation between concentrations and quantities is straightforward.

The dynamics of the system is ruled by two different reactions, namely *condensation* and *cleavage*. By means of the former two species are concatenated in order to create a longer species (e.g. $AB + BA \rightarrow ABBA$), whereas by means of the latter one specie is cut in order to create two shorter species (e.g. $ABBA \rightarrow AB + BA$ or $ABB + A$ or $A + BBA$), in general given a species of length l there are $2(l-1)$ different cleavage products.

We assume that for spontaneous reactions the rate of the backward reactions is negligible with respect to that of the forward reactions (i.e. strongly negative ΔG°).
Furthermore, since we are interested in the behaviors of catalytic reaction networks we assume that no reaction proceeds without the aid of catalysts, namely all reactions are characterized by a high energy barrier (i.e. activation energy) that would make them too slow to be observed in the absence of the correspondent catalysts.
The main novelty presented in this work is the explicit introduction of energy constraints, according to which some types of reaction require energy to proceed, as it will be described in the following section.
It is important to notice that the present version of the model neglects any catalysis provided by elements other than species belonging to the system, even though environmental catalysts, such as mineral clay, are thought to have played a relevant role in prebiotic synthesis (Ferris et al., 1996).

Given the number and the length of the species present in the system one can compute the overall number of conceivable reactions, including both cleavage and condensation, as

$$\hat{R} = \sum_{i=1}^{N} (L(x_i) - 1) + N^2 \,. \qquad (1)$$

where $L(x_i)$ is the length of the *i-th* species and N is the

total number of species present in the system. An important assumption is that we consider an independent probability p that any species catalyses a random reaction, hence not all the $\hat{R}$ conceivable reactions will occur, but only those that are actually catalyzed by some of the existing species. p turns out to be one of the key parameters of the model, since it rules the overall activity of the system by tuning the number of possible catalysts present in the reactor.

The dynamics is based on the well-known Gillespie stochastic algorithm (Gillespie, 1977, 2007) but, in order to speed up the computational performance, some of the processes are described by means of an approximated algorithm; in particular the ingoing and outgoing fluxes and, as we will see in the next section, the dynamics related to the introduction of the energy.

In accordance with the nature of the reactions, i.e. condensation and cleavage, we can summarize the reaction scheme as following:

- Cleavage: $AB + C \rightarrow A + B + C$
- Condensation: (whole reaction: $A + B + C \rightarrow AB + C$)
  - Complex formation: $A + C \rightarrow A : C$
  - Complex dissociation: $A : C \rightarrow A + C$
  - Final Condensation: $A : C + B \rightarrow AB + C$

where A and B are two generic substrates involved in a specific reaction, C is the specific catalyst for that reaction and $A : C$ represents a temporary complex, which is necessary for the condensation process to happen.

One of the main features of the model concerns the possibility to create new species by means of the internal dynamics. The creation of new species leads to the creation of new reactions; on the other side, some species could also vanish. To maintain the consistency of the system in the case of reappearance of some of the vanished species, all the reactions are kept in memory.

Another important remark regards the emergence of competition and inhibition phenomena by means of the particle-based algorithmic approach, since the molecules involved in a specific reaction cannot be used in another one at the same time.

Notice that with regard to an asynchronous stochastic model such this, the question on the correct reaction graph to use is of fundamental importance. To this end we introduced three distinct reaction graphs, to be used according to the circumstances. In detail:

- The *possible reactions graph*: in which all the possible reactions at a certain moment of the simulation, including those that will not actually occur, are drawn.
- The *complete reaction graph*: in which all the reactions that occur at least once over the simulation time frame are conserved.
- The *actual reaction graph*: after defining a specific temporal window, W, only the reactions that occur within W at a specific time are kept in the graph, while the older ones are removed. Notice that the temporal window turns out to be a key parameter in the analysis of the system, since the detection of ACSs is made using this specific graph representation. In this way it is possible to define cycles even in a stochastic system with asynchronous update and, at the same time, to neglect the influence of very rare reactions on the overall dynamics.

Introduction of the energy in the model

The first rationale at the base of the introduction of the energy within the model is that energy *does* exist in nature and, to a wide extent, it deeply affects the nature and the dynamics of biophysical and biochemical systems. With regard to our model of catalytic reaction network, both information and matter were present in the original description (Filisetti et al., 2010, 2011a), while energy was missing. Therefore, one of the major objectives of this work is to decipher whether and how energy actually influences the overall dynamical behavior. Moreover, we may hypothesize that the association of energy to some specific type of reaction could lead the system to novel complex behaviors, mostly in regard to the possible emergence of ACSs.

The general idea is to divide the possible reactions in two classes in accordance with the specific energetic constraints, namely *exergonic* and *endoergonic* reactions. While exergonic reactions occur releasing energy, endoergonic reactions require the presence of energy carrier molecules that release an amount of energy to some of elements involved in the reaction, otherwise the reaction cannot occur[2].
For simplicity we assume that the exergonic reactions release energy in form of heat (in the present state of the model there is no coupling between exergonic and endoergonic reactions) and that the presence of substrates and catalyst is sufficient for them to occur. Constraints on the endoergonic reactions are explained below.
It is also important to remark that, in order to maintain a certain degree of generality, no hypotheses on the specific form of energy are formulated. Temperature is assumed to be kept constant by coupling the reactions with a heat bath.

Let us assume the presence of an incoming flux of loaded energy carriers ϕ_E measured in (mol/sec), which transport

[2]We could assume the condensations to be endoergonic reactions and that, conversely, cleavage reactions occur spontaneously and do not require any chemical energy: these conditions hold, for example, in case of RNA in aqueous environment.

the energy into the system and which instantaneously diffuse in the reactor. The energy carriers, ECs from now on, bind and energize the internal species with a energization kinetic constant k_{nrg}.

Once that an energy carrier has released energy to a specific molecule it is removed from the system (we do not consider the unloaded energy carriers in the dynamics), while that species remains energized until it becomes part of any reactions requiring energy to proceed to completion. We also assume the presence of an outgoing flux of ECs coherent with the efflux constant of the reactor k_{out} and the presence of an ECs decay constant k_{dec}, by which an EC can be discarded because of the loss of its energetic load. Such processes are described as in the following:

$$\begin{cases} \frac{d[EC]}{dt} & = \phi_E - k_{out}[EC] - k_{dec}[EC] \\ & -k_{nrg}[EC][X^-] \\ \frac{d[X^+]}{dt} & = k_{nrg}[EC][X^-] - k_{out}[X^+] \\ & -k_{dec}[X^+] - \psi \\ \frac{d[X^-]}{dt} & = \psi + K + k_{dec}[X^+] - k_{nrg}[EC][X^-] \\ & -k_{out}[X^-] \end{cases} \quad (2)$$

where $[EC]$ stands for the concentration of the ECs, $[X^+]$ represents the overall concentration of the charged molecules, $[X^-]$ is the total concentration of the uncharged molecules, ψ represents the decrease of $[X^+]$, and the increase of $[X^-]$, because of the reactions occurred consuming the energy contained in the species involved in, and K represents the incoming flux (moles/sec) of uncharged molecules[3].

It is important to stress that, considering the three-molecular nature of the condensation reactions, and the bi-molecular nature of the cleavage reactions, there are 12 possible combinatorial energy configurations in accordance with the position of the molecules carrying the energetic group: catalyst and/or first substrate and/or second substrate, table 1.

In accordance with table 1, the reactions admitted by the possible different energy frameworks can be thought as two independent Boolean functions, one for the condensation reactions and one for the cleavage reactions, of the respectively 8 and 4 possible input arguments (there are 2^{2^k} possible different Boolean functions, where k is the number of different Boolean inputs).
Nevertheless, only a subset of the overall $256 + 16$ Boolean functions are biologically plausible according to the adopted assumptions.

[3] Although the model is based on a stochastic algorithm, in order to speed up the computational performance, both the energy flux and the species energization processes are described by means of an approximated algorithm.

	Catalyst	Substrate 1	Substrate 2
Condensation			
1	+	+	+
2	+	+	-
3	+	-	+
4	+	-	-
5	-	+	+
6	-	+	-
7	-	-	+
8	-	-	-
Cleavage			
9	+	+	//
10	+	-	//
11	-	+	//
12	-	-	//

Table 1: In the table all the possible energy configurations are represented. Symbol "+" stands for the charged state of the molecules whereas symbol "-" represents the uncharged state of the molecule.

In principle, if we consider a system composed of a set of distinct interacting chemicals, it would be reasonable to assign distinct energetic Boolean functions to each specific reaction. Of course, there are constraints between cleavage and condensation groups (for a nice and detailed presentation, see (Plasson and Bersini, 2009)); at the present stage of the model we make simple choices compliant with the underlying physical and chemical properties (see below), by leaving more detailed and complex scenarios to future developments.

Preliminary results

The preliminary analyses regarding the introduction of energy within the model are aimed to understand the influence of a variation of a) the energy carriers incoming flux ϕ_E and of b) the energization kinetic constant k_{nrg} on the overall dynamics, with particular attention to the emergence of ACSs.
In detail, we considered the specific case in which all the condensations are *endoergonic* reactions and, thus, require energy to occur, and all the cleavages are neutral, since they can occur both in presence and in absence of energy. Furthermore, we decided to concentrate on the case in which a unique Boolean energy function is associated to all the condensation reactions, i.e. the function number 14 (00001110 in binary code), which requires that at least one of the two substrates is energized, while the catalyst of the reaction is necessarily not energized. For the sake of completeness, the Boolean energy function associated to the cleavage is the number 15 (1111 in binary code), that is the *true* function.
We specify that, in the course of this study, we decided

to simulate systems with standard structural parameters[4] and with a *critical* reaction probability, i.e. the probability according to which one random species catalyses, on the average, one random reaction[5]. We made this choice in order to observe whether and to what extent an energy variation in the system affects the emergence of ACSs in the region of the parameters space that is, according to the literature (Farmer and Kauffman, 1986), close to the phase transition.

We analyzed different ensembles of systems in which we varied independently:

- the incoming flux of energy carriers ϕ_E, starting from the benchmark condition in which no carriers are present in the system: 0, 10^{-23} mol/sec (corresponding to 6 carriers/sec), 10^{-22} (60), 10^{-21} (600), 10^{-20} (6000);
- the energization kinetic constant k_{nrg}: 10^{-1}, 1, 10, 10^2, 10^3.

In figure 1 we can observe the variation of the average number of species present in the reactor, and not belonging to the incoming flux, at the end of the simulation (i.e. 1000 seconds) as a function of the variation of the incoming flux of energy carriers ϕ_E (*x axis*) and of k_{nrg} (*z axis*).

In those cases in which there are no energy carriers in the system we can see that no new new species are present at the end of the simulation and this is clearly due to the impossibility for the condensations to occur in case of no energy.
On the other hand, we can observe a maximum region along

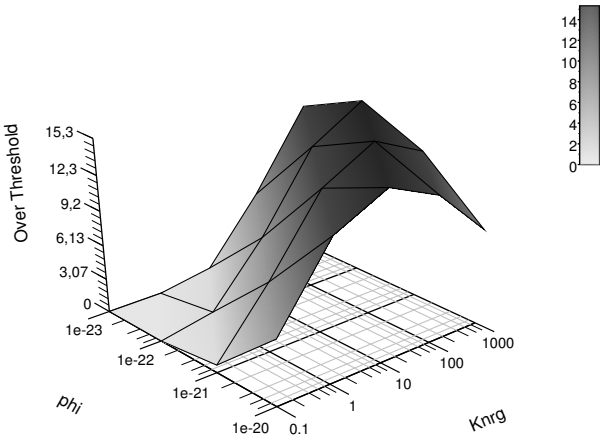

Figure 1: The figure shows the average number of species not belonging to the influx, with concentration greater than 0 from a set of 30 different simulations for each point represented in the graph. On the *x axis* the variation of ϕ_E is represented while on the *z axis* the variation of k_{nrg} is represented.

the direction of the diagonal individuated by the following combination of ϕ_E and k_{nrg} (respectively): (10^{-23} - 1; 10^{-22} - 10; 10^{-21} - 10^2; 10^{-20} - 10^3), the maximum of the slope being reached in the cases corresponding to the following three combinations of ϕ_E and k_{nrg} (respectively): (10^{-22} - 10; 10^{-21} - 10^2; 10^{-20} - 10^3), the first one being the absolute maximum.
Even if ϕ_E and k_{nrg} are independent parameters, their combination actually represents the amount of available energy present in the system: this is the reason why similar values of the variable under analysis (i.e. the number of new species) are observed in relation to different combinations of these parameters. Moreover, the presence of a region of maximum indicates that there is an optimal amount of energy for the system in terms of overall production of new species. For larger values of both ϕ_E and k_{nrg} the "efficiency" of the system in producing new species begins to decrease. This effect is partially due to the fact that when all the molecules in the reactor are energized the number of not-energized catalysts decrease due to the constraint on the total quantity of energy, as well as the number of possible condensations; hence in accordance with the particular assumptions concerning the chosen energy function, a decrement of the not-energized catalysts slows down the production of new molecules.
In figure 2 we can find the variation of the average concentration produced within the ACSs and within their first-order leaves in correspondence of the above mentioned combinations of ϕ_E and k_{nrg}.

In figure 3 the variation of the average concentration of

[4]The detailed setting of the system is the following:

- the alphabet is composed of two letters, A and B;
- the firing disk containing the elements present in the reactor at the beginning of the simulation is composed of all the species up to length 4;
- the volume of the reactor is set to 10^{-18} dm^3 and the overall initial concentration is set to 10^{-4};
- the influx is composed of all the species of the firing disk and the influx rate is set to 10^{-21} mol/sec;
- monomers and dimers can not be catalysts;
- the number of energy carriers entering the reactor and the value of the energy kinetic constant are varied according to the analyses and they are shown in the captions of the relative figures.

Notice that with these settings around 600 new molecules are entering the reactor every second and that at the theoretical dynamical equilibrium around 30000 molecules would be present inside the reactor.

[5]In this case the reaction probability is set to: 10^{-4}

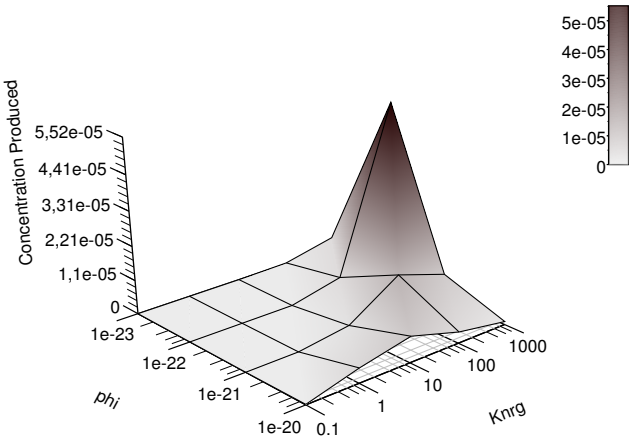

Figure 2: The figure shows the average concentration produced within the ACSs and within their first-order leaves from a set of 30 different simulations for each point represented in the graph. On the *x axis* the variation of ϕ_E is represented while on the *z axis* the variation of k_{nrg} is represented.

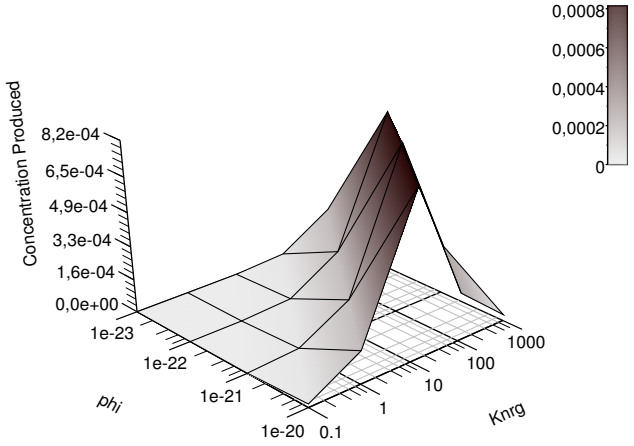

Figure 3: The figure show the average concentration produced by chains of reactions from a set of 30 different simulations for each point represented in the graph. On the *x axis* the variation of ϕ_E is represented while on the *z axis* the variation of k_{nrg} is represented.

the species produced by chains of reactions (and not belonging to ACSs) is shown.

We can see that, while the graph regarding the chains closely resembles that of the new species produced by the system, figure 1, confirming an optimal value of available energy in regard to the enhancement of the general activity, for what concerns the molecules produced within the ACSs (and their leaves) a unique point of maximum is reached for the combination $\phi_E = 10^{-22}$ and $k_{nrg} = 10^3$, which also corresponds to the maximum in the creation of new species. Finally, it is important to remark that with most of the combinations of ϕ and k_{nrg} no ACSs are present in the system at the end of the simulation and this would provide another possible explanation for the difficulty in observing the emergence of ACSs in wet lab experiments: according to these results, a fine tuning of the parameters regarding the energy is needed to allow the system to produce ACSs.

Conclusions

The introduction of energy constraints associated to specific types of reactions represents a major novelty in the development of our stochastic model of catalytic reaction networks. In this regard, the main aim of this work was to show whether and to what extent the introduction of energy might affect the overall dynamics and, in particular, the emergence of autocatalytic cycles.

To this end, the preliminary analyses on critical systems showed that the combination of two key parameters, namely the incoming flux of energy carriers ϕ and the energization kinetic constant k_{nrg}, jointly representing the amount of energy available for the endoergonic reactions, is responsible for a remarkable variation of the general activity of the system (indirectly attested by the production of new species). In particular, it was possible to prove the existence of an optimal value of energy, beyond which the activity of the system begins to decrease.

Focusing on the emergence of ACSs, it was then possible to demonstrate that the maximum production of new species is observed in the case of systems with optimal values of energy, which contain ACSs involving a large number of molecules, hence confirming their relevance in the overall dynamics. On the other hand, with most of the tested combinations of ϕ and k_{nrg} ACSs could not been detected and this might provide another possible explanation to the difficulty in observing their emergence in wet lab experiments. Moreover, as we already showed in (Filisetti et al., 2011a, 2010), the autocatalytic sets are not robust and in most of them the catalytic closure is achieved by means of a "bottleneck" reaction occurring rarely during the simulations; although the energy constraints allow the emergence of structural ACSs, they do not confer neither robustness nor some forms of self-sustaining dynamics.

The results show that our model might unravel some unexpected features concerning the emergence of autocatalytic sets of molecules as for example the presence of an optimum in the energy flux.
Thus, several developments are underway in order to refine the description of the model like, for instance, the associa-

tion of distinct energization Boolean functions and of distinct k_{nrg} to different reactions and species, with the purpose of investigating possible complex behaviors related to the availability of energy.

Acknowledgements

This work has been partially supported by the Fondazione di Venezia, http://www.fondazionedivenezia.it, (DICE project). Useful discussions with Stuart Kauffman, Norman Packard and Ruedi Füchslin are gratefully acknowledged.

References

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., and Walter, P. (2005). *Molecular Biology of the Cell, Fifth Edition*. Garland Science, 4 edition.

Anastasi, C., Buchet, F. F., Crowe, M. A., Parkes, A. L., Powner, M. W., Smith, J. M., and Sutherland, J. D. (2007). Rna: prebiotic product, or biotic invention? *Chemistry & biodiversity*, 4(4):721–39.

Bagley, R. and Farmer, J. D. (1992). Spontaneous emergence of a metabolism. *Artificial Life II. Santa Fe Institute Studies in the Sciences of Complexity*, X:93–141.

Bagley, R. J., Farmer, J. D., Kauffman, S. A., Packard, N. H., Perelson, A. S., and Stadnyk, I. M. (1989). Modeling adaptive biological systems. *Bio Systems*, 23(2-3):113–37; discussion 138.

Bartel, D. P. (1999). *Re-creating an RNA Replicase*, chapter 5, pages 1431–144.

Bartel, D. P. and Unrau, P. J. (1999). Constructing an rna world. *Trends in cell biology*, 9(12):M9–M13.

Budin, I. and Szostak, J. W. (2010). Expanding roles for diverse physical phenomena during the origin of life. *Annual review of biophysics*, 39:245–63.

De Lucrezia, D., Anella, F., and Chiarabelli, C. (2007). Question 5: on the chemical reality of the rna world. *Orig Life Evol Biosph*, 37(4-5):379–385.

Dyson, F. J. (1985). *Origins of life*. Cambridge: Cambridge University Press.

Eigen, M. and Schuster, P. (1977). The hypercycle. a principle of natural self-organization. part a: Emergence of the hypercycle. *Die Naturwissenschaften*, 64(11):541–65.

Farmer, J. and Kauffman, S. (1986). Autocatalytic replication of polymers. *Physica D: Nonlinear Phenomena*, 220:50–67.

Ferris, J. P., Hill, A. R., Liu, R., and Orgel, L. E. (1996). Synthesis of long prebiotic oligomers on mineral surfaces. *Nature*, 381(6577):59–61.

Filisetti, A., Graudenzi, A., Serra, R., Villani, M., De Lucrezia, D., Fuchslin, R. M., Kauffman, S. A., Packard, N., and Poli, I. (2011a). A stochastic model of the emergence of autocatalytic cycles. *Journal of Systems Chemistry (jn press)*.

Filisetti, A., Serra, R., Villani, M., Füchslin, R. M., Packard, N. H., Kauffman, S. A., and Poli, I. (2010). A stochastic model of autocatalytic reaction networks. In *Proceedings of the European Conference on Complex Systems (ECCS)*, Lisbon, September 13-17.

Filisetti, A., Serra, R., Villani, M., Graudenzi, A., Füchslin, R. M., and Poli, I. (2011b). The influence of the residence time on the dynamics of catalytic reaction networks. *Frontiers in Artificial Intelligence and Applications - Neural Nets WIRN10 - Proceedings of the 20th Italian Workshop on Neural Nets*, pages 243–251.

Fishkis, M. (2007). Steps towards the formation of a protocell: the possible role of short peptides. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life*, 37(6):537–53.

Fuchslin, R. M., Filisetti, A., Serra, R., Villani, M., DeLucrezia, D., and Poli, I. (2010). Dynamical stability of autocatalytic sets. In Fellermann, H., Dörr, M., Hanczyc, M. M., Laursen, L. L., Maurer, S., Merkle, D., Monnard, P.-A., Stoy, K., and Rasmussen, S., editors, *Artificial Life XII, Proceedings of the Twelfth International Conference on the Synthesis and Simulation of Living Systems*, pages 65–72. The MIT Press.

Gilbert, W. (1986). Origin of life: The rna world. *Nature*, 319(6055):618–618.

Gillespie, D. T. (1977). Exact stochastic simulation of coupled chemical reactions. *The Journal of Physical Chemistry*, 81(25):2340–2361.

Gillespie, D. T. (2007). Stochastic simulation of chemical kinetics. *Annual Review of Physical Chemistry*, 58(1):35–55.

Jain, S. and Krishna, S. (1998). Autocatalytic set and the growth of complexity in an evolutionary model. *Phys Rev Lett*, 81:5684–5687.

Kaneko, K. (2006). *Life: An Introduction to Complex Systems Biology (Understanding Complex Systems)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA.

Kauffman, S. A. (1986). Autocatalytic sets of proteins. *J Theor Biol*, 119(1):1–24.

Kauffman, S. A. (1993). *The Origins of Order: Self-Organization and Selection in Evolution*. Oxford University Press, USA, 1 edition.

Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K., and Ghadri, M. R. (1996). A self-replicating peptide. *Nature*, 382(6591):525–8.

Lifson, S. (1997). On the crucial stages in the origin of animate matter. *Journal of molecular evolution*, 44(1):1–8.

Ma, W., Yu, C., Zhang, W., Zhou, P., and Hu, J. (2010). Self-replication: spelling it out in a chemical background. *Theory in biosciences*.

Miller, S. L. (1953). A production of amino acids under possible primitive earth conditions. *Science*, 117:528–529.

Mossel, E. and Steel, M. (2005). Random biochemical networks: the probability of self-sustaining autocatalysis. *Journal of theoretical biology*, 233(3):327–36.

Müller, U. F. (2006). Re-creating an rna world. *Cellular and molecular life sciences : CMLS*, 63(11):1278–93.

Oparin, A. (1924). *The origin of life on the earth*. Oliver and Boyd, 1st ed., p edition.

Plasson, R. and Bersini, H. (2009). Energetic and entropic analysis of mirror symmetry breaking processes in a recycled microreversible chemical system. *The journal of physical chemistry. B*, 113(11):3477–90.

Rios, A. C. and Tor, Y. (2009). Model systems: how chemical biologists study rna. *Current opinion in chemical biology*, 13(5-6):660–8.

Ruiz-Mirazo, K. and Mavelli, F. (2008). On the way towards 'basic autonomous agents': stochastic simulations of minimal lipid-peptide cells. *Bio Systems*, 91(2):374–87.

Saghatelian, A., Yokobayashi, Y., Soltani, K., and Ghadiri, M. R. (2001). A chiroselective peptide replicator. *Nature*, 409(6822):797–801.

Segre, D., Lancet, D., Kedem, O., and Pilpel, Y. (1998). Graded autocatalysis replication domain (gard): kinetic analysis of self-replication in mutually catalytic sets. *Orig Life Evol Biosph*, 28(4-6):501–514.

Talini, G., Gallori, E., and Maurel, M.-C. (2009). Natural and unnatural ribozymes: back to the primordial rna world. *Research in microbiology*, 160(7):457–65.