

# The role of backward reactions in a stochastic model of catalytic reaction networks

Alessandro Filisetti<sup>1,4</sup>, Alex Graudenzi<sup>2</sup>, Chiara Damiani<sup>2</sup>, Marco Villani<sup>1,3</sup> and Roberto Serra<sup>1,3</sup>

<sup>1</sup>European Centre for Living Technology, University Ca' Foscari of Venice, Italy

<sup>2</sup>Dept. of Informatics, Systems and Communication, University of Milan Bicocca, Italy

<sup>3</sup>Dept. of Physics, Informatics and Mathematics, Modena and Reggio Emilia University

<sup>4</sup>CIRI, Energy and Environment Interdepartmental Center for Industrial Research, University of Bologna

corresponding author: alex.graudenzi@unimib.it

## Abstract

We investigate the role of backward reactions in a stochastic model of catalytic reaction network, with specific regard to the influence on the emergence of autocatalytic sets (ACSs), which are supposed to be one of the pre-requisites in the transition between non-living to living matter.

In particular, we analyse the impact that a variation in the kinetic rates of forward and backward reactions may have on the overall dynamics.

Significant effects are indeed observed, provided that the intensity of backward reactions is sufficiently high. In spite of an invariant activity of the system in terms of production of new species, as backward reactions are intensified, the emergence of ACSs becomes more likely and an increase in their number, as well as in the proportion of species belonging to them, is observed. Furthermore, ACSs appear to be more robust to fluctuations than in the usual settings with no backward reaction.

This outcome may rely not only on the higher average connectivity of the reaction graph, but also on the distinguishing property of backward reactions of recreating the substrates of the corresponding forward reactions.

## Introduction

Models of catalytic reaction networks have been widely investigated in the last decades, with different goals and purposes, yet mostly in regard to the broad theme of the origin of life and with the design of artificial *protocells* (Carletti et al., 2008; Filisetti et al., 2010; Rasmussen et al., 2004; Serra et al., 2007; Szostak et al., 2001).

In particular, in the quest for a reasonable theory describing the transition from non-living to living matter, many frameworks have been proposed, among others the *metabolic-first* scenario (Dyson, 1985; Smith and Morowitz, 2004; Wächtershäuser, 1990; de Duve, 1982), the *protein-first* hypothesis (Oparin, 1924; Fox, 1974; Lee et al., 1996, 1997; Issac and Chmielewski, 2002), the *compartmentalization* (Bachmann et al., 1992), the *compositional* approach (Segré et al., 1999; Segre et al., 1998; Segré and Lancet, 2000) and the *gene-first* hypothesis included in the RNA world theory (Gilbert, 1986; Müller, 2006; De Luca et al., 2007; Anastasi et al., 2007; Talini et al.,

2009; Rios and Tor, 2009; Budin and Szostak, 2010). Even if the dispute is far from being concluded (Cornish-Bowden and Cárdenas, 2008; Stano and Luisi, 2010; Schrum et al., 2010; Budin and Szostak, 2010), one of the underlying key requirements in most of these theories is that the production of the molecular species involved in the transition relies on robust reaction pathways.

In this regard, some theories account for linear chemical pathways capable of producing the sufficient amount of species at energy-rich sites, e.g. hydrothermal vents (Ogasawara et al., 2000) or under plausible prebiotic conditions (Costanzo et al., 2009). Nevertheless, in most of the cases the emergence of sets of collectively self-replicating molecules, i.e. autocatalytic cycles (or *autocatalytic sets*, ACSs from now on)<sup>1</sup> appears to be an essential requirement to achieve the self-sustenance and the evolvability of the system. Indeed, there are many examples of ACSs in current biological systems, which are the outcome of billions of years of evolution. Therefore, the investigation of the generic properties of catalytic reaction networks, with particular respect to the sufficient conditions for the emergence of ACSs and the characterisation of their dynamical properties, is fundamental<sup>2</sup>.

<sup>1</sup>A classical definition of ACS is that of a subset of chemicals in which the production of each element is catalysed by at least another elements belonging to the subset (Kauffman, 1986). Hereinafter, a more formal definition with specific regard to our model will be provided.

<sup>2</sup>It is very important to remark that the presence of ACSs only is not sufficient to define *life*, which it is largely believed to require also the presence of a container that separates the living system from the environment, as well as a coupling between the replication rate of the internal molecules and the growth and division rate of the container. This theme is at the centre of the research on protocells.

In previous works (Serra et al., 2007; Filisetti et al., 2008; Carletti et al., 2008) we proved that, once that such a coupling is achieved, the rates of the replication of the internal molecules and that of the growth of the container tend to spontaneously synchronise through successive divisions. This also leads to an exponential growth of the population of protocells that, in turns, implies a Darwinian selection process among them (Munteanu et al., 2006).

Furthermore, in Serra et al. (2013) we introduced the first known

To this end, different models have been proposed by, e.g., Dyson (Dyson, 1985), Eigen and Schuster (Eigen and Schuster, 1977; Eigen and McCaskill, 1988; Eigen and Schuster, 1978), Kauffman (Kauffman, 1986; Hordijk et al., 2010; Hordijk and Steel, 2004; Mossel and Steel, 2005), Jain and Krishna (Jain and Krishna, 1998), Lancet (Segre et al., 1998; Segré and Lancet, 2000), Kaneko (Kaneko, 2006) and Vasas and Fernando (Vasas et al., 2012). Despite the theoretical differences, most of the models predict a phase transition leading to the spontaneous emergences of ACSs, after matching certain key conditions, either structural or dynamical according to the cases. Note that, given that the presence of ACSs may eventually lead the system to display remarkable discrepancies between the concentration of the involved molecules from that of analogous systems with no ACSs, this could be actually investigated through wet-lab experiments.

Nonetheless, it is indeed difficult to detect the emergence of ACSs in lab experiments and this could be due, on the one hand, to the somehow drastic simplifications at the basis of the theoretical models or, on the other hand, to the fact that real experiments never matched the requirements suggested by the theories.

In order to fill the gap between theories and experiments and provide insights for further experimentation, in Filisetti et al. (2011c) we introduced a novel model of catalytic reaction network, based on a fully stochastic framework and in which the system complexity can grow according to the dynamics, through the creation of new species and reactions. The model takes inspiration from the works by Kauffman (Kauffman, 1986) and its subsequent developments by others (Bagley et al., 1989; Bagley and Farmer, 1992; Farmer and Kauffman, 1986; J.D.Farmer et al., 1986).

The model considers abstracted entities accounting for monomers and polymers (i.e. *species*) and simplified interactions among them, in terms of cleavages (i.e. the cutting of two species) and condensations (i.e. the concatenation of two species). The key constraint of the model is that each reaction must be catalysed in order to occur. In this regard, any species in the system can be selected to be the catalyst of any possible reaction with a certain probability. The system's dynamics is then stochastically simulated within an open flow reactor. By using a stochastic framework, it is possible to consider in an adequate way the relevance of noise, random fluctuations and low-numbers-effects on the overall dynamics, most of all when dealing with systems

---

stochastic model in which a catalytic reaction network is modelled within a simplified model of protocell. Although a stochastic description has been adopted also in Mavelli and Ruiz-Mirazo (2013), our model of protocell deals with the capability to create new molecular species by means of the reactions present in the system.

close to the phase transition in which the emergence of ACSs becomes plausible.

We remark that the focus of the model is *not* on the detailed characterisation of the entities and reactions of a specific chemical system, but rather on the investigation of the dynamical behaviour that emerges from the interaction of simple entities, with the final goal of deciphering the generic (or universal) properties, that is, those that are shared by a possibly broad range of different chemical systems. In particular, we aim at determining the minimal conditions for the emergence of ACSs and the sensitivity of the phenomenon to variations in some key parameters.

In this regard, in our previous works we studied in depth the influence that variations in some of the key parameters of the system has on the overall dynamical behaviour and on the production of ACSs.

One first result we obtained was to detect that a variation in the composition of the set of molecules present at the beginning of the simulation does not seem to remarkably affect the dynamics of the systems, whereas modifications in the incoming flux seem to deeply influence the overall behaviour (Filisetti et al., 2011c).

For this reason, we focused our attention on the incoming flux composition and diversity (Filisetti et al., 2011a). The results of the analysis that we performed showed that the a variation in the number of distinct species belonging to the incoming flux influences the general activity of the system: considering a fixed overall incoming flux concentration, the larger the number of diverse species (regardless their lower individual concentration), the higher the activity of the system, in terms of overall number of species and molecules, yielding a larger number of ACSs. On the contrary, the length of the polymers belonging to the flux seems not to be so relevant.

Another key parameter of the system, the average residence time of the molecules within the reactor, was also analysed, suggesting that the larger the residence time is, the higher the probability of emergence of ACSs is.

In another work, presented at *ECAL 2011* (Filisetti et al., 2011b), we introduced some plausible energy constraints associated to specific types of reactions, to investigate whether and how the introduction of a form of energy could affect the dynamics and the emergence of ACSs. Preliminary analyses showed that there exists an optimal combination of two key parameters, i.e. the incoming flux of energy carriers and the energization kinetic constant (which account for the amount of energy available for endoergonic reactions) and that this combination ensures a larger production of new species. Further research is needed for a better understanding of the phenomenon.

Finally, one of the most important results was to highlight the general fragility of the ACSs that have been observed. In fact, their existence usually depends on same rare molecules

and reactions, which may disappear because of random fluctuations, hence preventing the autocatalytic closure over a significant span of time. This outcome could provide one of the first possible explanations of the difficulty in observing the emergence of ACSs in wet-lab experiments.

In line with this methodological approach, within this work we carry on with the analysis of the key parameters of the system, in order to provide a more complete and coherent picture of the phenomenon.

In particular, we here relax one of the constraints of the classical formulation of the model, that is, the exclusion of backward reactions. In previous studies backward reactions were neglected, by hypothesising that the Gibbs energy  $\Delta G$  for any reaction is large enough to maintain the system far from the chemical equilibrium. Considering that backward reactions *do* occur in nature, here we want to investigate their role in the overall dynamics.

We define a cleavage (respectively, a condensation) as the *backward reaction* of a specific condensation (respectively, cleavage), i.e. the *forward reaction*, if:

- its products are the substrates of that condensation (respectively, cleavage),
- its substrates are the products of that condensation (respectively, cleavage).

The analysis has then been focused on the effects of the relative strength of the rates of the forward and the backward reactions.

In section II the model will be briefly outlined. In section III results of the simulations with backward reactions will be shown. Finally, in section IV the discussion and some indications for future works will be presented.

## The model

A detailed description of the model can be found in (Filisetti et al., 2011a,c), here we will only outline the key features.

The model represents an open system in which monomer and polymers, i.e. the *species*, are involved in catalysed reactions. Every species  $x_i, i = 1, 2, \dots, N$  is defined by an ordered string of letters selected from an arbitrary alphabet (e.g. A, B, C...) and its amount, either concentration or quantity, i.e. number of *molecules*. The only allowed reaction types are: i) *cleavage*, the splitting of two species (e.g.  $AAAB \rightarrow A + AAB$ ) and the ii) *condensation*, i.e. the concatenation of two species (e.g.  $BBAA + BA \rightarrow BBAABA$ ), which requires an intermediate step involving the formation of a temporary complex between the substrate and the catalyst.

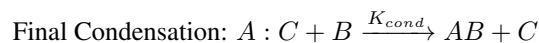
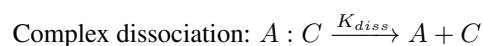
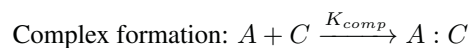
We neglect spontaneous reactions by assuming that there is a sufficiently high activation energy for any reaction scheme. Therefore, only catalysed reactions are allowed and every

species  $x_i$  (longer than a specific threshold) can be selected to be the catalyst of a given reaction with a certain (uniform) probability  $p_i = p, i = 1, 2, \dots, N$ . Therefore, the reaction scheme is defined in a probabilistic way, i.e. in different simulations the same species can be the catalyst of distinct reactions. Besides, the initial reaction scheme can dynamically evolve and increase in dimension because of the creation of new species, which can be (probabilistically) involved in either novel reactions as substrates, products or catalysts, provided that the coherence with the existent reaction scheme is maintained. The set of couples  $\{species, reaction\}$  in which the species catalyses the reaction defines the *chemistry* of the system, because it describes a coherent possible artificial world. Hence, it is possible to simulate distinct chemistries or to keep the chemistry fixed and simulate different time histories.

In the classical formulation of the model, backward reactions are also excluded, by hypothesising that the Gibbs energy  $\Delta G$  for any reaction is large. The main goal of this work is to investigate the implications of relaxing this constraint.

A possible example of each reaction type is shown:

- Cleavage:  $AB + C \xrightarrow{K_{cleav}} A + B + C$
- Condensation: (whole reaction:  $A + B + C \rightarrow AB + C$ )



A and B are two random species standing for the substrates of the specific reaction, C is the catalyst of that reaction and  $A : C$  is the transient complex.  $K_{cleav}$ ,  $K_{comp}$ ,  $K_{diss}$  and  $K_{cond}$  respectively are the kinetic rates of cleavage, complex formation, complex dissociation and final condensation<sup>3</sup>. The outgoing flux is simulated by assigning a common decay time  $K_{out}$  to each species and complex. The incoming flux rate  $K_{in}$  is measured in moles per second and the average residence time is given by  $1/K_{out}$ .

The dynamics of the system is simulated through the well-known Gillespie algorithm (Gillespie, 1977) for the stochastic simulation of chemical reaction system, with the key modification of allowing the creation of new species and reactions that are not present in the initial conditions. In particular, the system is modelled within a *continuous stirred-tank reactor* (CSTR), which allows continuous

<sup>3</sup>Notice that a parameter sensitivity analysis of the model was presented in Damiani et al. (2013). The main goal was to identify those kinetic parameters that mainly influence the ability of the system to increase the diversity of the species.

ingoing and outgoing fluxes of molecules. In another work (Serra et al., 2013) we introduced a semi-permeable membrane to separate the catalytic reaction network from the environment.

Two possible representations of the system are possible, which results in different graphs. The first concerns the catalytic activity of the system: an edge from  $x_c$  to  $x_i$  is drawn if species  $x_c$  is the catalyst of the reaction in which the species  $x_i$  is one of the products. This leads to the so-called *catalyst-product* graph. The second representation regards the assembling activity: an edge from  $x_j$  to  $x_i$  is drawn when  $x_j$  is a substrate involved in a reaction in which  $x_i$  is one of the products. This allows to draw the *substrate-product* graph.

Besides, the adoption of an asynchronous stochastic framework implies the problem of detecting cycles in a univocal way. We lastly decided to introduce a graph in which every edge (either catalyst-product or product-substrate) is maintained only if the specific reaction occurs within an arbitrary temporal window,  $W$ . We call it the *actual reaction graph* and can be applied to both the catalyst-product and the substrate-product graphs. In this way, the influence of very rare reactions is neglected and cycles can be coherently detected.

In particular, in the context of our model, we define as ACS a subset of species which belong to a *strongly connected component* (SCC) in the catalyst-product actual reaction graph<sup>4</sup>.

**The introduction of backward reactions.** The goal of this work is to investigate whether and how the introduction of backward reactions may influence the overall dynamics of the system and, in particular, with respect to the emergence of ACSs.

To this end, we define a backward reaction for any existing reaction of the system (which will be defined as forward reaction), relative to both cleavages and condensations (see the definition of forward and backward reactions in the previous section). In the example scheme above, the condensation is the backward reaction of the cleavage (or the other way round).

The analysis is then focused on the variation of a key parameter  $R$ , which accounts for the relationship between the forward and the backward reactions kinetic rates, and is defined as follows. Note that, given that in the current configuration of the system we set  $K_{comp} = K_{cond}$  for all the condensations, only  $K_{cond}$  will be included in the definition.

We distinguish two cases:

- if the forward reaction is a cleavage, given any  $K_{cleav}$ :

$$R = \frac{2K_{cleav}}{K_{cond}} \rightarrow K_{cond} = \frac{2K_{cleav}}{R} \quad (1)$$

- if the forward reaction is a condensation, given any  $K_{cond}$ :

$$R = \frac{K_{cond}}{2K_{cleav}} \rightarrow K_{cleav} = \frac{K_{cond}}{2R} \quad (2)$$

Varying  $R$  it is possible to define different ratios between the rates of forward and backward reactions and, accordingly, given the kinetic rates of any forward reaction to determine those of the corresponding backward reaction<sup>5</sup>.

## The simulations

The benchmark for this kind of analysis is the case in which no backward reactions are considered and that will be indicated with *NOREV* from now on. We then considered 4 values of  $R = 1, 10, 100, 1000$ .

We created 10 different chemistries and for each of them we varied the value of  $R$  only (simulating 10 different histories, for a total of 500 distinct simulations), in order to disentangle the effect of its variation on the dynamics. The details of the simulations can be found in the caption of Fig. 1.

In Fig. 1 we display the (average) number of distinct species present in at least one copy in time. No remarkable differences are detectable in the number of different species, which reaches an asymptotic value around 60 after a transient whose length is around 300 seconds, in all the cases. Notice that the number of distinct species (which somehow accounts for the *diversity* of the system) does not depend on the flux dynamics only, but on the general capability of the system of generating new species and reactions. Hence, this outcome suggests that the overall activity of the system, in terms of production of new species, is not enhanced by the introduction of backward reactions.

Moreover, a relatively moderate variation is observed also with regard to the asymptotic total number of molecules in the system, which is around 30.000 for all the distinct cases (not shown here).

The (average) number of molecules and that of species belonging to ACSs are shown in Fig. 2. We here have a indeed remarkable result: in correspondence of the lowest values of  $R$  (i.e. proportionally faster backward reactions) we observe a clear increase of the percentage of both molecules and species belonging to ACS, starting from the

<sup>5</sup>Notice that the factor 2 in Eq. 1 and 2 was chosen, in accordance with our previous works (Fuchslin et al., 2010), to roughly balance the speed of the cleavage, which is a one-step reaction, with that of the condensation, which is a 2-steps reaction.

<sup>4</sup>See the footnote number 1 at page 1 for a more general definition

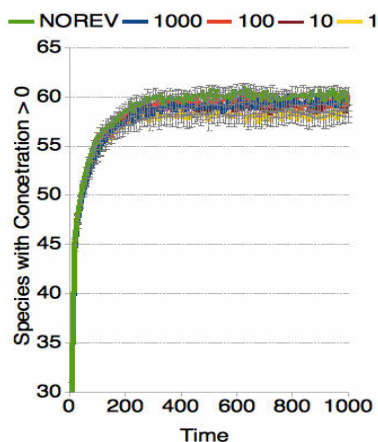


Figure 1: Variation of the average number of species present in the system in time. The five lines represent the case with no backward reactions (*NOREV*) and those of systems characterised by values of  $R = 1, 10, 100$  and  $1000$ . The x axis represents time (in seconds). The values are averaged over 100 different simulations for each value of  $R$ . The bars represent the standard error.

The settings of the simulations are the following.

Alphabet: [A,B]; probability of catalysis,  $p$ : 0.00097; volume of the reactor:  $1e^{-18}L$ ; overall concentration =  $1e^{-4}M$ ; set of species in the influx: all the species up to length 4; minimum polymer length to have catalytic activity: 2; baseline  $K_{cond} = 50M^{-1}sec^{-1}$ , baseline  $K_{comp} = 50M^{-1}sec^{-1}$ , baseline  $K_{cleav} = 25M^{-1}sec^{-1}$ , baseline  $K_{diss} = 1e^{-6}sec^{-1}$ ; influx rate =  $1e^{-21}mol/sec$ ; simulation time: 1000 sec. 10 different chemistries are created, for each chemistry 5 different systems are created: with no backward reactions, with  $R = 1, 10, 100$  and  $1000$ ; for each system 10 different histories are simulated. The number of simulations is so 500.

benchmark case of no backward reactions, in which less than 5% of the molecules and of the species belong to ACSs (and analogously for slower backward reactions, i.e.  $R = 100$  and  $1000$ ), then the case of  $R = 10$ , in which around 20% of the molecules and of the species are in ACSs and up to case in which  $R = 1$ , which involves around 40% of the molecules and 45% of the species in the ACS dynamics. This result hints at a very important consideration: the faster the backward reactions are<sup>6</sup>, the more likely the emergence of ACSs with a large number of molecules and species is. It is even possible to hypothesise a threshold in  $R$  after which the emergence of large ACS becomes very likely, which would be, in this case, between  $R = 10$  and  $R = 100$ .

<sup>6</sup>Meaning proportionally faster with respect to forward reactions.

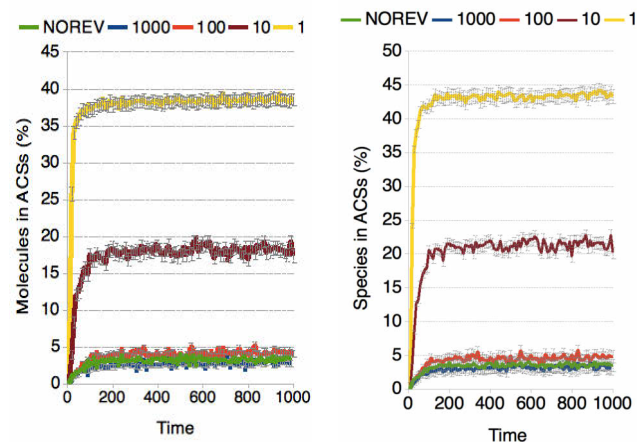


Figure 2: Variation of the average percentage of molecules (left) and species (right) belonging to ACSs in time, with respect to the cases: *NOREV*,  $R = 1, 10, 100$  and  $100$ . The x axis represents time (in second). The bars display the standard error. The percentage is computed by looking at the molecules and species present at any time step in the system.

In Fig. 3 we report the variation of the number of ACSs (left) and that of percentage of species belonging to ACSs (right) in time, with respect to all the different simulations. Each row of the graph represents a distinct simulation, so it is possible to follow the dynamical evolution of any simulated system, with regard to these two key variables. By looking at the left graphs, regarding the number of ACSs in time, one first important result proves what stated above by analysing the average values. In correspondence of lower values of  $R$  a larger number of simulations is characterised by: i) the emergence of at least one (usually robust in time) ACS, ii) a larger number of distinct ACSs. Whereas for systems with no or very slow backward reactions (e.g.  $R = 1000$ ) in many case no ACSs emerge, when ACSs emerge are often not persistent in time (showing an oscillatory fashion) and the maximum number of observed ACSs is around 4, for low values of  $R$  ( $R = 1$  or  $R = 10$ ) in almost all the simulations at least one ACS is observed and we even observe simulations which yield a indeed large number of ACSs (up to 10 for  $R = 1$ ).  $R = 100$  seems to characterise an intermediate condition, perhaps close to a phase transition in which the emergence of ACSs becomes indeed likely.

Besides, given that the simulations are ordered in sets of 10 different histories for each one of the 10 chemistries, it is possible to notice how some chemistries are actually more efficient in producing ACSs, by looking at the large clearer stripes (i.e. 10 rows, corresponding to 10 histories of the same chemistry), e.g. for  $R = 10$  or  $R = 1$ .

In the right panels it is possible to observe the percentage of species belonging to ACSs in every simulation. For the cases in which only one or a few ACSs emerge (no

backward reactions or high  $R$ ) the fraction of species belonging to ACSs seems to show a strong correlation with the number of ACSs itself. Nevertheless, in the cases in which a larger number of ACSs emerge (low values of  $R$ ) we notice that this correlation is not always preserved and there are, on the one hand, simulations in which a very large percentage of species belong to a relatively low number of (big) ACSs and, on the other hand, others in which a large number of (small) ACSs involve a relatively low proportion of the species. This outcome would suggest that systems with relatively faster backward reactions display more heterogeneous dynamical behaviours.

It is also important to remark that while in the case with no or slow backward reaction at most 20/25% of the species are involved in the ACS dynamics, for low values of  $R$  this percentage increases dramatically, up to more than 70% for some simulation with  $R = 1$ , in which the dynamics of the system is monopolised by ACSs.

In Fig. 4 we show the percentage of molecules in ACSs at the end of the simulation (i.e. time = 1000) as a function of the average connectivity of the catalyst-product actual reaction graph (at time 1000) for all the different cases.

One can first notice that the systems with lower  $R$  correspond to higher average connectivities. This is a somehow expected result, given that the introduction of backward reactions implies an increase in the number of possible reactions and also that the lower the value of  $R$  is, the higher the probability of occurrence of these reactions is, resulting in a actual reaction graph which is increasingly more connected.

Besides, by looking at the line that interpolates the dots relative to the distinct cases (from the case of no backward reaction to the case of  $R = 1$ ), one can detect an apparently super-linear trend, which may underly some non-linear phenomenon, possibly related to the intrinsic nature of backward reactions. In fact, we here remark that the introduction of backward reactions does not simply imply a larger average connectivity for the system, as one of the features of backward reactions is to continuously recreate (as products) the substrates of the corresponding forward reactions. In particular, this action ensures the maintenance of the chains of reactions that guarantee the continuous flow of materials from the system's input toward the ACS structures. It is unlikely that the same reinforcement action of the ACSs' sustaining chains is provided simply because of the doubling of the number of reactions. In order to avoid the collapse, in fact, each ACS has to exactly guarantee the presence of the materials it is consuming: randomly created reactions have scarce chances to reinforce all the needed chemical species, whereas backwards reactions are automatically pointing toward the correct substrates. Given the autocatalytic nature of the ACS, this action is supposed

to guarantee the presence of the needed catalysts<sup>7</sup>.

Therefore, even though further analyses are needed to address this issue, we may suppose that this phenomenon entails important implications on the dynamics and stability of ACSs and, accordingly, to the percentage of molecules belonging to them.

## Conclusions and further developments

In this work we investigated the role of backward reactions in a stochastic model of catalytic reaction network in an open reactor.

The introduction of backward reactions involves significant changes in the overall dynamics, with particular regard to the emergence of ACSs, provided that their speed (hence, frequency of occurrence) is sufficiently high, as established by the kinetic rates and, in particular, by the proportion between the kinetic rates of forward and backward reactions. In other words, the intensity of backward reactions is fundamental to observe remarkable differences in the overall dynamics.

In detail, despite an observed substantial invariance in the number of different species (i.e. the diversity of the system) produced by the dynamics, as long as the relative values of the rates for the forward and backward reaction are decreased (i.e. the intensity of backward reactions is increased), an always higher number of these species is involved in an always larger number of different ACSs, in an increasing number of different simulations.

Besides, when backward reactions gain intensity the ACSs appear to be also more robust to variations and oscillations in time. This could represent a very significant result, mostly in regard with the dynamical fragility of ACSs that was observed in our previous analyses of systems without backward reactions.

It is also possible to hypothesize the presence of a threshold above which the likelihood of emergence of resistant ACSs dramatically increases.

One partial explanation of this general outcome is that backward reactions indeed add new reactions to the chemistry, leading to an increase of the average connectivity of the reaction network, which has been considered one of the key variables in regard to the emergence of ACSs (Filisetti et al., 2011c; Farmer and Kauffman, 1986; Jain and Krishna, 1998).

Nonetheless, the key property of backward reactions of recreating the substrates of the relative forward reactions could be essential in influencing the process of emergence of ACSs, not only because of the increase of the number of possible reactions, but mostly because of their action of reinforcement in favor of the supply chains supporting the

<sup>7</sup>Notice that we are currently designing experiments in order to separate the effects of doubling the number of reactions from those deriving by simple enabling the backward reactions.

existence of ACS structures. Note that this reinforcement action is stronger when the values of the reactions' forward and backward kinetic constants are closer, a circumstance that may have deeply influenced the chemical composition of the first historically functioning ACS structures. Analyses underway are aimed at addressing this issue. Furthermore, the phenomenon of competition for the same catalyst between forward and backward reactions could be another interesting phenomenon to investigate.

In another work (Serra et al., 2013) we introduced a model of catalytic reaction network in protocell, by considering the simplest possible architecture, that is a semi-permeable membrane that selects the species that can enter or exit the protocell.

Among the various results it was shown that protocells display distinct asymptotic behaviours, according to different variables, a property that has never been observed in CSTRs.

Preliminary analyses on the introduction of backward reactions in the protocell model would suggest that even mildly intense backward reactions would lead the system toward a more homogeneous dynamical behaviour. Even if further investigations are ongoing, this outcome would suggest another interesting role of backward reactions in this kind of system, also hinting at possible differences in the hypothesised threshold on the proportion among the kinetic rates.

### Acknowledgements

We deeply thank Stuart Kauffman for many inspirational discussions.

### References

- 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.
- Bachmann, P. A., Luisi, P. L., and Lang, J. (1992). Autocatalytic self-replicating micelles as models for prebiotic structures. *Nature*, 357(6373):57–59.
- 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.
- 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.
- Carletti, T., Serra, R., Villani, M., Poli, I., and Filisetti, A. (2008). Sufficient conditions for emergent synchronization in protocell models. *J Theor Biol*, 254(4):741–751.
- Cornish-Bowden, A. and Cárdenas, M. L. (2008). Self-organization at the origin of life. *Journal of theoretical biology*, 252(3):411–8.
- Costanzo, G., Pino, S., Ciciriello, F., and Di Mauro, E. (2009). Generation of long RNA chains in water. *The Journal of biological chemistry*, 284(48):33206–16.
- Damiani, C., Filisetti, A., Graudenzi, A., and Lecca, P. (2013). Parameter sensitivity analysis of stochastic models: Application to catalytic reaction networks. *Computational biology and chemistry*, 42:5–17.
- de Duve, C. (1982). PEROXISOMES AND RELATED PARTICLES IN HISTORICAL PERSPECTIVE. *Annals of the New York Academy of Sciences*, 386(1 Peroxisomes a):1–4.
- 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 Mccaskill, J. (1988). Molecular Quasi-Specie. *J Phys Chem*, 81:6881–6891.
- 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.
- Eigen, M. and Schuster, P. (1978). The Hypercycle: a Principle of Natural Self-Organisation, Part B. *Naturwissenschaften*, 65(7):7–41.
- Farmer, J. and Kauffman, S. (1986). Autocatalytic replication of polymers. *Physica D: Nonlinear Phenomena*, 220:50–67.
- 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*, 2(1):2.
- Filisetti, A., Graudenzi, A., Serra, R., Villani, M., De Lucrezia, D., and Poli, I. (2011b). The role of energy in a stochastic model of the emergence of autocatalytic sets. In T, L., M, G., Bersini, H., Bourguine, P., Dorigo, M., and Doursat, R., editors, *Advances in Artificial Life, ECAL 2011 Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems*, pages 227–234. MIT Press, Cambridge, MA.
- Filisetti, A., Graudenzi, A., Serra, R., Villani, M., Fuchslin, R. M., Packard, N., Kauffman, S. a., and Poli, I. (2011c). A stochastic model of autocatalytic reaction networks. *Theory in biosciences = Theorie in den Biowissenschaften*, pages 1–9.
- Filisetti, A., Serra, R., Carletti, T., Villani, M., and Poli, I. (2008). Synchronization phenomena in protocell models. *Biophysical Reviews and Letters (BRL)*, 3(1/2):325–342.
- Filisetti, A., Serra, R., Carletti, T., Villani, M., and Poli, I. (2010). Non-linear protocell models: synchronization and chaos. *The European Physical Journal B*, 77(2):249–256.
- Fox, S. W. (1974). The Protein Theory of the Origin of Life. *American Biology Teacher*, 36:161–172.

- 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.
- Hordijk, W., Hein, J., and Steel, M. (2010). Autocatalytic Sets and the Origin of Life. *Entropy*, 12(7):1733–1742.
- Hordijk, W. and Steel, M. (2004). Detecting autocatalytic, self-sustaining sets in chemical reaction systems. *Journal of theoretical biology*, 227(4):451–461.
- Issac, R. and Chmielewski, J. (2002). Approaching Exponential Growth with a Self-Replicating Peptide. *Journal of the American Chemical Society*, 124(24):6808–6809.
- Jain, S. and Krishna, S. (1998). Autocatalytic set and the growth of complexity in an evolutionary model. *Phys Rev Lett*, 81:5684–5687.
- J.D.Farmer, Packard, N. H., and Perelson, A. S. (1986). The immune system, adaptation, and machine learning. *Physica D*, 22(2):187–204.
- 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.
- 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.
- Lee, D. H., Severin, K., Yokobayashi, Y., and Ghadiri, M. R. (1997). Emergence of symbiosis in peptide self-replication through a hypercyclic network. *Nature*, 390(6660):591–4.
- Mavelli, F. and Ruiz-Mirazo, K. (2013). Theoretical conditions for the stationary reproduction of model protocells. *Integrative Biology*, pages 324–341.
- 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.
- Munteanu, A., Attolini, C. S., Rasmussen, S., Ziock, H., and Solé, R. V. (2006). Generic Darwinian selection in protocell assemblies. DOI: SFI-WP 06-09-032, *SFI Working Papers*, Santa Fe Institute.
- Ogasawara, H., Yoshida, A., Imai, E., Honda, H., Hatori, K., and Matsuno, K. (2000). Synthesizing oligomers from monomeric nucleotides in simulated hydrothermal environments. *Origins of life and evolution of the biosphere : the journal of the International Society for the Study of the Origin of Life*, 30(6):519–26.
- Oparin, A. (1924). *The origin of life on the earth*. Oliver and Boyd, 1st ed., p edition.
- Rasmussen, S., Chen, L., Deamer, D., C.Krakauer, D., Packard, N. H., Stadler, P. F., and Bedau, M. A. (2004). Transitions from Nonliving to Living Matter. *Science*, 303, 963-965.
- Rios, A. C. and Tor, Y. (2009). Model systems: how chemical biologists study RNA. *Current opinion in chemical biology*, 13(5-6):660–8.
- Schrum, J. P., Zhu, T. F., and Szostak, J. W. (2010). The Origins of Cellular Life. *Cold Spring Harbor perspectives in biology*, 2(9):a002212.
- Segré, D., Ben-eli, D., Deamer, D. W., and Lancet, D. (1999). The lipid world. *Origins of Life and Evolution of the Biosphere*, 31(1):119–145.
- Segré, D. and Lancet, D. (2000). Composing life. *EMBO reports*, 1(3):217–22.
- 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.
- Serra, R., Carletti, T., and Poli, I. (2007). Synchronization phenomena in surface-reaction models of protocells. *Artificial life*, 123(2):123–38.
- Serra, R., Filisetti, A., Villani, M., Damiani, C., Graudenzi, A., and Panini, T. (2013). A stochastic model of catalytic reaction networks in protocells. *Submitted to Artificial Life*.
- Smith, E. and Morowitz, H. J. (2004). Universality in intermediary metabolism. *Proceedings of the National Academy of Sciences of the United States of America*, 101(36):13168–73.
- Stano, P. and Luisi, P. L. (2010). Achievements and open questions in the self-reproduction of vesicles and synthetic minimal cells. *Chemical communications (Cambridge, England)*, 46(21):3639–53.
- Szostak, J. W., Bartel, D. P., and Luisi, P. L. (2001). Synthesizing life. *Nature*, 409(6818):387–390.
- 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.
- Vasas, V., Fernando, C., Santos, M., Kauffman, S., and Szathmari, E. (2012). Evolution before genes. *Biology Direct*, 7(1):1.
- Wächtershäuser, G. (1990). Evolution of the first metabolic cycles. *Proceedings of the National Academy of Sciences of the United States of America*, 87(1):200–4.



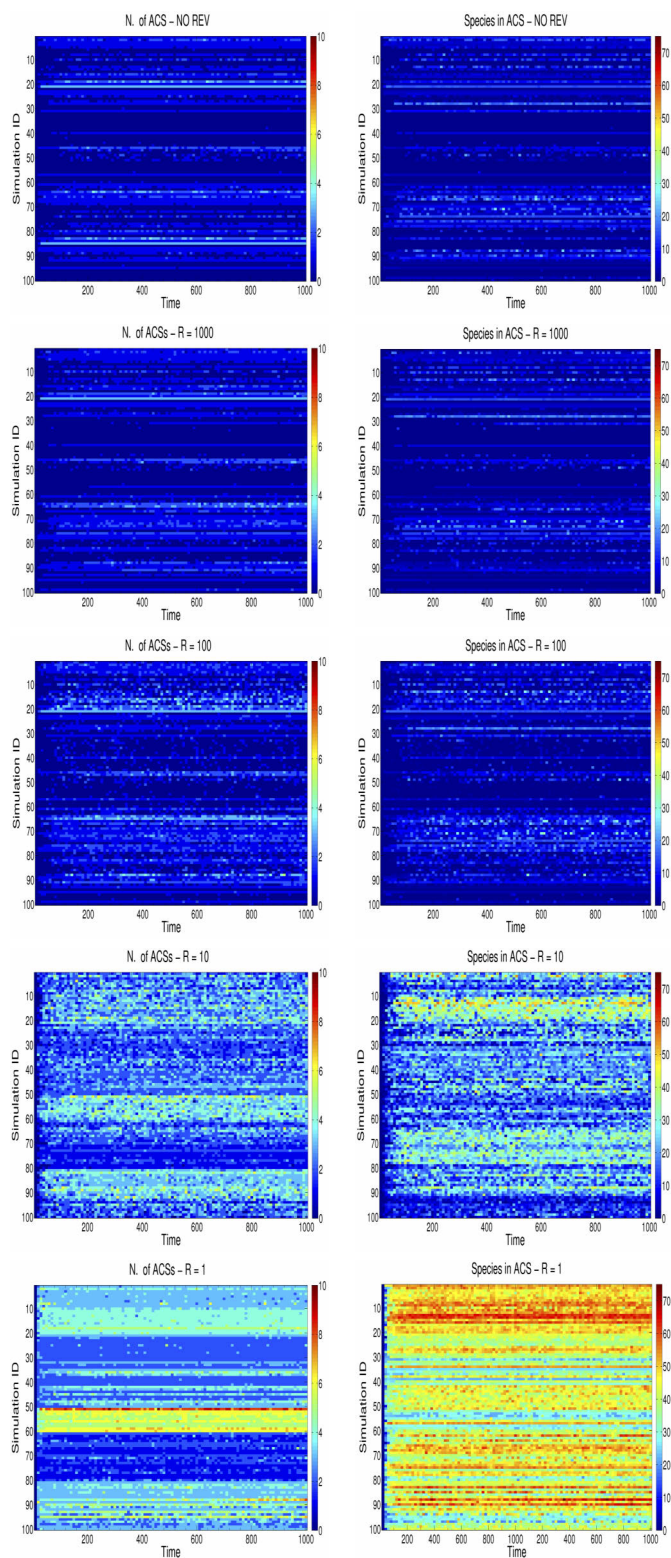


Figure 3: Variation of the number of ACSs (left) and of the percentage of species (on the total) belonging to ACSs (right) for all the simulations, with respect to the cases (in order from top to bottom): *NOREV*,  $R = 1000$ , 100, 10 and 1. Each row represents one distinct simulation, the x axis represents time. The colours stand for the value of that variable, as in the corresponding colour legend.

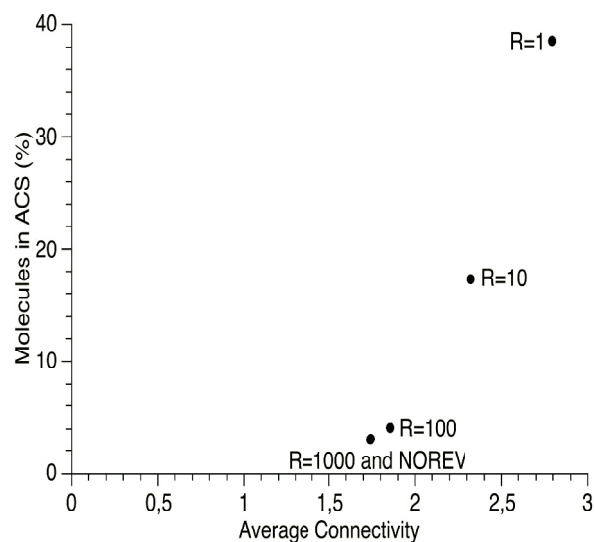


Figure 4: Phase diagram of the relation between the average connectivity of the system and the average percentage of molecules belonging to ACSs at the end of the simulation (i.e. time = 1000 seconds), with respect to the cases: *NOREV*,  $R = 1$ , 10, 100 and 100. The x axis stand for the average connectivity and the y axis for the percentage of molecules in ACSs. The colors represent the different cases.