On RAF sets and autocatalytic cycles in random reaction networks

Alessandro Filisetti[1], Marco Villani[2,1], Chiara Damiani[3,6], Alex Graudenzi[3], Andrea Roli[4], Wim Hordijk[5], and Roberto Serra[2,1]

[1] European Centre for Living Technology, University Ca' Foscari of Venice, Italy
[2] University of Modena and Reggio Emilia, Department of Physics, Informatics and Mathematics
[3] University of Milano Bicocca, Department of Informatics, Systems and Communication
[4] University of Bologna, Dept. of Computer Science and Engineering (DISI)
[5] SmartAnalytiX.com, Lausanne, Switzerland
[6] SYSBIO - Centre for Systems Biology, Piazza della Scienza 2, 20126 Milano, Italy

Abstract. The emergence of autocatalytic sets of molecules seems to have played an important role in the origin of life context. Although the possibility to reproduce this emergence in laboratory has received considerable attention, this is still far from being achieved.

In order to unravel some key properties enabling the emergence of structures potentially able to sustain their own existence and growth, in this work we investigate the probability to observe them in ensembles of random catalytic reaction networks characterized by different structural properties.

From the point of view of network topology, an autocatalytic set have been defined either in term of strongly connected components (SCCs) or as reflexively autocatalytic and food-generated sets (RAFs).

We observe that the average level of catalysis differently affects the probability to observe a SCC or a RAF, highlighting the existence of a region where the former can be observed, whereas the latter cannot. This parameter also affects the composition of the RAF, which can be further characterized into linear structures, autocatalysis or SCCs.

Interestingly, we show that the different network topology (uniform as opposed to power-law catalysis systems) does not have a significantly divergent impact on SCCs and RAFs appearance, whereas the proportion between cleavages and condensations seems instead to play a role.

A major factor that limits the probability of RAF appearance and that may explain some of the difficulties encountered in laboratory seems to be the presence of molecules which can accumulate without being substrate or catalyst of any reaction.

1 Introduction

The dynamics of sets of interacting molecular species, in those cases where new molecular types can be created by the interactions among some of the existing

ones, pose formidable problems. This study is of the outmost importance in researches on the origin of life (OoL) but its applications might be interesting also within laboratories or industrial applications.

When only few new molecular types can be generated, the complications are manageable. The actual interesting case is the one where many types might appear. In order to study the generic features of this kind of systems, the use of random molecular species has been introduced [14, 10, 12, 11], in particular in the context of the so-called binary polymer model [21, 9, 22] where all the species are linear strings made out of a binary alphabet. Since catalysis plays a prominent role in biological processes, it is often also assumed that only catalyzed reactions take place at an appreciable rate. In the spirit of searching for generic properties, randomness is also extensively used to determine catalysts: every molecular species has a certain fixed probability to catalyze a certain reaction.

In this way it is possible to represent the reactions as a directed graph, where there is an edge from a molecular species to all those species whose production reactions are catalyzed by that species. To complete the picture, it is worth stressing that only two types of reactions are allowed, condensation (which creates a new string by concatenating two previous ones) and cleavage (which generates two strings by separating two parts of a preexisting one). The model is described in more details in section 2.

The graph described above links catalysts to the products of the reactions they catalyze; it can therefore be considered as a *catalyst → product* graph. The products may in turn catalyze other reactions. It has been often suggested that the presence of cycles (similar to those that are found in present-day biological systems [1, 6–9, 20, 10, 12, 29]) plays a key role in the dynamics of these systems, since cycles of catalysts give rise to a collectively autocatalytic system. To be precise, we will call *SCC* a Strongly Connected Component [5] of the network described above, i.e. a subset of the entire network where each node is reachable (directly or indirectly) from every other node in the subset. A *SCC* is therefore composed of one or more molecular types where the formation of each member is catalyzed by at least another member of the *SCC*.

While some models (e.g. [19]) suppose that the presence of a catalyst is sufficient for a reaction to occur, in the binary polymer model considered here, in order for cleavage or condensation reactions to occur, the substrates must be present as well. These substrates may belong to what is called the food set F, which involves for example those molecules of the feed those are supposed to be freely available or to be continuously supplied from outside; these molecules might or might not have catalytic properties. Substrates that are not in the food set may be present too, i.e. molecular species that take part in reactions and are produced and consumed by the system itself.

The importance of the substrates is properly captured by the notion of Reflexive Autocatalytic and Food generated (RAF) sets, that is a set of molecular types and reactions where each member can be generated by other members of the set through a series of catalyzed reactions, starting from F [16, 18, 17]. Note that the identification of this structure needs knowledge about all the chemical

species (substrates, products, catalysts) involved in the reaction scheme: a more detailed description can be found in section 2.

SCCs and RAFs are indeed important tools to analyze and understand the structure of sets of interacting molecules: note that an SCC may not be a RAF, given that substrates are not taken into account, and that a RAF may also not be an SCC, given that cyclic structures are not strictly required (a RAF may turn out not to have cycles if at least one of the species of the food set is also a catalyst).

Here below we analyze how frequent SCCs and RAFs are as a function of the average level of catalysis $\langle c \rangle$ (the average number of reactions catalyzed by each chemical species). Note that the behavior of the model turns out to be trivial when $\langle c \rangle$ is very small (no SCCs and no RAFs are observed) or very high (the system is full of both). Instead, it is interesting to consider an intermediate region. An outcome of the present work is that there exists a region of $\langle c \rangle$, similar results have observed in [17] observing both RA and RAF sets, that we term the gap region, where SCCs can be observed, but RAFs are not (these sentences are to be interpreted in a statistical sense, see the results in section 3 for the details). Therefore, in this region, the very existence of cycles in the catalyst $\rightarrow$ product networks does not imply that they actually affect in a sensible way the main network properties.

In the present work we will investigate protocell systems, compartments containing sets of molecules able to collectively self-replicate, able to undergo fission and proliferate [34, 26]. Several protocell architectures have been proposed [30, 4, 13, 15, 23, 24, 27, 32, 33], most of them identify the compartment with a lipid vesicle that may spontaneously fission under suitable circumstances. In the specific, we are here interested on a very simple protocell model, where a semi-permeable membrane embrace a small physical space where all the key reactions take place in the aqueous phase, and it allows the passage of small molecules only [31].

In the present work we investigate sets of autocatalytic reactions where a constant concentration of some small chemical species is guaranteed: this situation can be maintained by chemo-physical systems as protocells (compartments containing sets of molecules able to collectively self-replicate, able to undergo fission and proliferate [34, 26, 31]—an interesting system for both Origin of Life and technological themes) or CSTRs (Continuous-flow Stirred-Tank Reactor [28]).

Thus, the topological structures having the highest probabilities to emerge in such systems, by taking a purely graph theoretical approach, will be investigated, as in previous papers by Kauffman, Steel and Hordijk [17, 18], and the existence of the gap region will be shown. It is however tempting to speculate about the behavior of a truly dynamical model in this region.

We have already shown in our previous works [14, 12, 11, 10] that, for $\langle c \rangle$ values close to the threshold for SCCs appearance, those that are observed tend to be fragile, some of their links refer to very rare reactions and do not really affect the behavior of the system [12, 10]. We guess that this behavior is related to the absence of RAFs, and that by exploring the region close to the second threshold (i.e. the one for RAFs) one should be able to find more productive net-

works. Moreover, we noted that (within the size of the systems we explored) the thresholds of SCCs and RAFs appearance in uniform and power-law catalysis systems are very similar. Interestingly, RAF systems at different average levels of catalysis are composed of different mixtures of autocatalysis, SCCs and linear structures. Different proportions of cleavages and condensations lead to different probabilities of RAFs appearance, and the presence of a large fraction of reactions producing long molecules (condensations ligating long chemical species) could inhibit the RAFs appearance as we will be discussed in section 3. So, the RAF presence is influenced in different ways by different characteristics of the chemical system.

In section 3 the main results are presented, while the discussion of further work is deferred to the final section 4.

2 Description of the model and network topologies

A detailed description of the theoretical model can be found in Kauffman [22, 21] and in Filisetti et al. [12, 10]. In the following only the main features, useful for the scope of this work, will be summarized. Random catalytic reaction networks are composed of a set of species $S = (s_1, s_2, \cdots, s_n)$. In particular, species are concatenation of characters from left to right taken from a finite alphabet, in this work a binary alphabet $A = \{1; 0\}$. Thus, the cardinality $|S|$ of S is $|S| = 2^{M+1} - 2$, where M is the maximum length of the system. In accordance with the original version of the model [21], each $s_i \in S$ can take part to two types of reaction only: condensation, where two species are concatenated to create a longer species (e.g. $100 + 10 \rightarrow 10010$), and cleavage, where a species is divided in two shorter species (e.g. $1001 \rightarrow 10 + 01$); s_i can be either a substrate or a product according to the position within the reaction scheme. Each species has an independent probability $p_{s_i} = 1/|S|$ to be selected as a substrate for a reaction, hence in case of condensation two species will be selected while in case of cleavage just one species will be selected. The overall number of conceivable reactions $\hat{R}$, i.e., all the reaction schemes allowed by the combination of all the possible $|S|$ molecular species, varies according to the constructional method adopted; in this work we typically adopt the method described in the following.[7]

Reactions are created so that the product of the reaction cannot be longer than M, hence the number of conceivable reactions is equal to:

$$\hat{R} = 2 \cdot \sum_{i=2}^{M} 2^i \cdot (i - 1) \tag{1}$$

where the multiplicative term 2 indicates that each reaction scheme is in principle valid both for cleavage reactions and condensation reactions.[8] Not all the reactions belonging to $\hat{R}$ are catalyzed. So, it is possible to compute an average

---

[7] The only exception is presented in the final part of the results section.

[8] Forward and backward reactions, during the creation of the reactions graph, are in principle handled as two separated reactions.

level of catalysis $\langle c \rangle = R/|S|$ which denotes the average number of reactions catalyzed by each single species.[9] Therefore, each catalysis can be represented as a pair (r_k, s_i) where $r_k \in \hat{R}$ stands for the selected reaction and $s_i \in S$ behaves as a catalyst for r_k.

The procedure of catalysis assignation leads to different catalytic reaction network topologies. In this work, two kinds of topologies will be assessed, based on uniform and preferential assignation. In the former case, each $s_i \in S$ has the same probability $p_k^i = 1/|S|$ to be a catalyst for a whatever reaction k, while in the latter case p_k^i is determined by the chemical characteristics of the involved species. In real systems, researchers observed that very few species can catalyze many reactions, whereas many species catalyze only one reaction, or none. This situation can be represented by means of a catalysis distribution having the shape of a power law, our second topology.[10]

The entire set $\mathcal{C} = (r_k, s_j)$ of catalysis form what we call "artificial chemistry", hence different random instantiations of $\mathcal{C}$ lead to different random catalytic reaction networks, i.e., different "artificial chemistries"; in figure 1 an example of a complete reaction graph is depicted while in figure 2 the catalyst $\rightarrow$ product representation of the graph shown in figure 1 is reported.

The structural role of backward reactions is evaluated as well. Thus, artificial chemistries imposing the presence of both forward and reverse reactions will be created.[11] It is worthwhile to notice that introducing reverse reactions leads to a double number of reactions catalyzed, so in case of reverse reactions the number of reactions to catalyze is divided by 2, $\hat{R} \rightarrow \hat{R}/2$.

Analysis will be carried out measuring the emergence of reflexively autocatalytic sets (RAF sets) on the complete graph $CRN = (S, R, \mathcal{C})$, where $\mathcal{C}$ represents catalysis, and investigating the strongly connected components emerging in the graph representation considering catalysts and products only, hence showing the catalytic activity of the system without taking care of the presence and the nature of the substrates. In figure 1 two examples of RAF sets and two SCCs are shown. In particular, the group of molecular species and reactions belonging to the group named "$RAF + SCC$" form a SCC and a RAF at the same time.

---

[9] Since different M will be assessed, in this work we prefer to adopt the average level of catalysis $\langle c \rangle$, as in [16], instead of the standard reaction probability p, as we adopted in our previous works [10, 12], and Kauffman [21] and others [2, 25] in important works on this topic. Nevertheless, according to $|S|$, it is always possible to move from $\langle c \rangle$ to p and *vice-versa*.

[10] We obtain the power law distribution by slightly modifying the algorithm proposed by Barabási and Albert [3]. We increase p_k^i as a function of the already catalyzed reactions, i.e., the probability to catalyze a reaction is weighted with the number of reactions already catalyzed so that $p_k^i = \#r_i / \sum_{z=1}^{|S|} \#r_z$, where $\#r_i$ and $\#r_z$ indicate the number of reactions already catalyzed by the i-th and the z-th species respectively. In such a way, we obtain a power-law distribution in the number of reactions each chemical species can catalyze.

[11] Since the analysis are static and only statistic structural properties of the networks will be assessed, the choice on which reaction is the direct one and which is the reverse one is only implementative and does not affect the analysis.
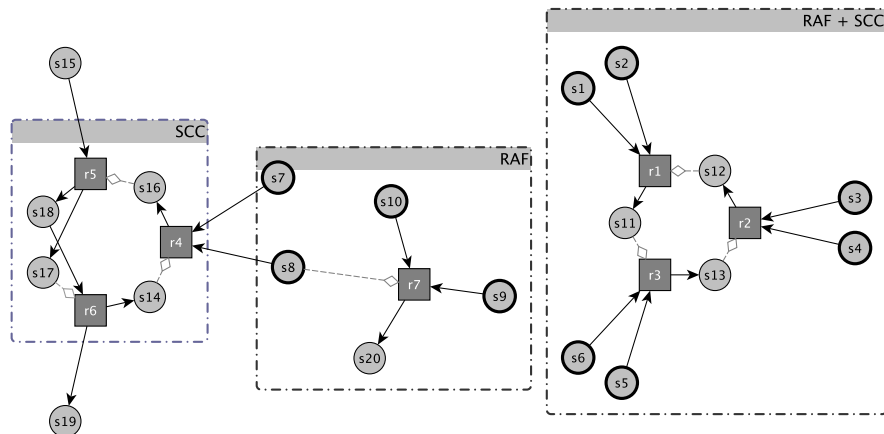
Fig. 1. The picture shows an example of representation of a catalytic reaction network by means of a complete reaction graph, taking into account both molecular species and reactions. Circles stand for molecular species and in particular bold circles represent species belonging to the food set F. Squares depict reactions, straight arrows indicate the participation of a species to the reaction as substrate (edge points to the reaction) or product (edge starts from the reaction). Catalysis are represented by dotted gray arrows. For the sake of comprehension, both molecular species and reactions forming RAFs and SCCs are grouped together.

On the opposite, the other two reaction structures are, from left to right, a SCC only and a RAF set only.

3 Results

As discussed above, we investigate here catalytic reaction networks, where only small molecular species have a constant concentration (this situation can be maintained by protocells [31], or CSTR systems [12]). In the following, scenarios where the length of the buffered species (the "foodset" of the from which RAF structures are built) ranges from 2 ($|S| = 6$) to 4 ($|S| = 62$) will be considered.

We are interested now in analyzing the role of different scenarios on both the presence and the internal structure of RAFs within protocells. The assessed scenarios involve different distributions of the chemicals' catalyzing capabilities (a uniform distribution, and a slightly more realistic power-law distribution, which allows the presence of a little number of highly versatile catalysts), the presence or absence of backward reactions, and the possibility for long chemical species to be not suitable for participating in any reaction, leading in such a way to the accumulation of useless chemical species (for example, because lack of solubility or folding difficulties).
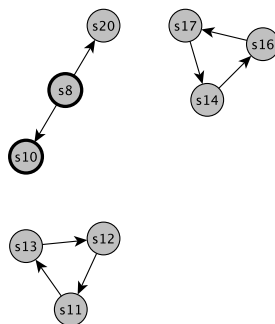
Fig. 2. The picture shows the catalyst → product representation of the complete graph $\mathcal{C}$ shown in figure 2. In this representation just the catalytic activity of the reaction network is depicted, and no information about the substrate of the reactions are present. Circles stand for molecular species and in particular bold circles represent species belonging to the food set F. In this case straight arrows indicate the catalysis of a particular molecular species by means of its catalyst. By means of this representation, the two SCCs are clearly visible and they are formed from s_11, s_12, s_13 and s_14, s_16, s_17 respectively.

3.1 Strongly Connected Components and RAFs

The first observation regards the presence of SCCs and RAFs within the reaction graphs in accordance with a foodset composed of species up to length 2 only, i.e., when the membrane allows the transfer of chemicals having maximum length 2. Figure 3 shows the fraction of network instances showing at least 1 SCC (1 RAF), by varying $\langle c \rangle$ and M. It can be observed that the SCCs and RAFs transition zones significantly differ (the higher the maximum length of the system, the higher the transition slope): the vast region between 1.0 and 2.5 levels of catalysis is therefore rich of SCC structures unable to self-sustain because the failed fulfilling the closure condition. [12]

The showed case regards scale-free topologies without backward reactions (obviously, in these experiments we maintain constant the total number of reactions). It is worth mentioning that the bias induced on the system topology by the pairing of forward-backward reactions do not change the positions of the transition zones nor their slopes (data not shown here), as long as condensation and cleavages are equally present - we will discuss the effect of uneven presence of condensation and cleavages in the next paragraph. Remarkably, almost iden-

---

[12] It is important to notice that in [17] the transition is found at 1.25. Such a difference is basically related to a different way to count forward and reverse reactions. Hordijk [17] consider both forward and backward reactions as a unique reaction whereas in our work, though it is clear that they account for the same reaction scheme, we consider them as two different reactions.

tical behaviors hold for random topologies having the same number of chemical species and reactions (data not shown, and [17]).[13]
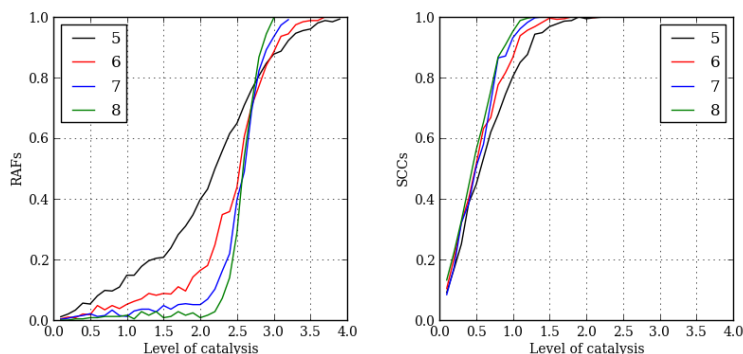
Fig. 3. The fraction of simulations showing at least 1 RAF (left) and 1 SCC (right), by varying $\langle c \rangle$ and M in networks with forward reactions only. F contains all the species up to length 2. On the x-axis the average level of catalysis $\langle c \rangle$ is represented while on the y-axis the fraction of network instances (out of 1000 networks for each $\langle c \rangle$) is depicted.

3.2 The inner RAF structure

The aforementioned results are derived from systems with protocell membranes that allow the diffusion of chemical species up to length 2 ("BUF_2" scenario). Since F does not show any catalytic activity, in that case all RAFs are necessarily composed of at least one SCC. On the contrary, if we extend the membrane semi-permeability to longer species, some of these objects may be able to catalyze some chemical reactions: this is a crucial change, because the formation of RAFs composed of linear structures (on the catalyst $\rightarrow$ products graph) turn out to be possible, thus catalytic roots are buffered ("BUF_3" scenario).

This change in F composition has an apparently huge effect on the presence of RAF, which appears also at low level of catalysis (figure 4); note however that the difference among chemistries with different maximum lengths M ("MaxLen M" systems) consists mainly in lower and lower probabilities of choosing the buffered species as root of linear structures on the catalyst $\rightarrow$ products graph

---

[13] In [17], the authors show the so-called RA sets—which are reflexively autocatalytic but not necessarily food-generated—whose transition and nature correspond to that of SCC, where the transitions happen on the same zones of scale-free topologies in case of non catalytic activity of the foodset molecular species, i.e, "BUF_2" scenario in this work.
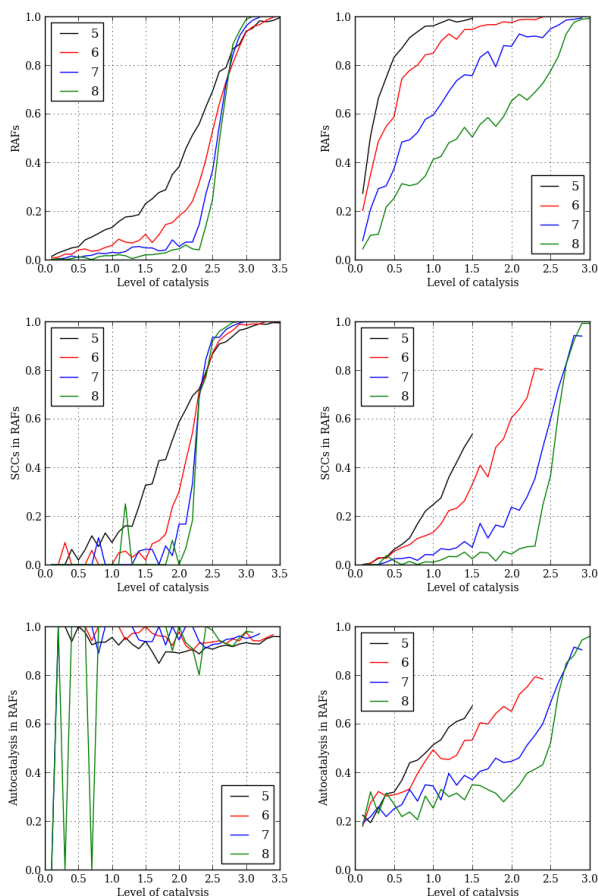
Fig. 4. TOP: Fraction of simulations showing at least 1 RAF. MIDDLE: fraction of RAFs having at least 1 SCC. BOTTOM: fraction of RAFs having at least 1 autocatalysis. On the x-axis $\langle c \rangle$ is represented, F is composed of all the species up to length 2 (left panel) and 3 (right panel). For each $\langle c \rangle$ and for each value of M, 1000 network instances have been created. For computational reasons, once the 100% of networks with a specific $\langle c \rangle$ contain at least a RAF set, the system automatically goes to the next M, thus in some cases the analysis on SCC does not reach a $\langle c \rangle = 4$, the maximum level of catalysis evaluated.

(a more interesting use of the "MaxLen M" scenarios is presented in the next paragraph). Indeed, the presence of the 2.5 threshold can be grasped on the final part of "MaxLen 8" scenarios: it is a clue that the line showing greatest chemistries (where there is a very low probability of choosing small molecules as catalysts) tends towards the plot of "BUF_2" scenario (where small molecules

are not catalysts). Such ample systems (and the "BUF_2" scenario as well) may represent a good approximation of real chemistries.

More interesting differences are observed when comparing the fraction of systems having at least one SCC or one autocatalysis within the obtained RAFs. [14] At low catalysis levels, within the "BUF_2" scenario, almost all RAFs contain an autocatalysis, whereas the formation of other SCC structures inside the RAF is unlikely; on the contrary, as the average catalysis level grows up, the fraction of RAFs entailing an SSC tends to 1 (slightly before the 2.5 zone) - note that these RAFs still have an high probability of containing an autocatalysis as well. The situation within the "BUF_3" scenario differs substantially: the sum of SCCs and autocatalysis do not reach the 100%, and this gap increases as the maximum allowed length decreases, that is, the prevailing structures for a large zone of catalysis level are linear chains on the catalyst $\rightarrow$ product graph. SCCs and autocatalysis play a minor role. [15]

Within a scenario that allows the diffusion of chemical species up to length 4 (data not shown here), the growth in the probability to observe a RAF happens even earlier than in the "BUF_3" scenario, whereas the presence of SCCs and autocatalysis within the RAF is almost identical between the two cases (indeed, the probability of having SCCs and autocatalysis are identical, not depending on the number of the buffered chemical species). So, in "BUF_4" scenario the presence of linear chains on the catalyst $\rightarrow$ product graph is even more significant. Remarkably, similar graphs and values hold for the scale-free topologies we tested (same number of chemical species and reactions).

3.3 Cleavages, condensations and irreversibility

Chemical reactions are generally reversible, in the sense that the reaction proceeds both forward (from reactants to products) and backward (in the opposite direction); however, note that, when far from equilibrium, the rates of the two processes are not equal. Particular environments (as aqueous phases rather than oleic phases) can hugely influence the reaction rates of all cleavages and/or of all condensations, whereas the presence or absence of catalysts can hugely influence the reaction rates of some specific reactions (for instance the particular sets of compartments, chemical species and reactions present inside the current living systems can hugely influence the reaction rates inside cells).

Hence, in case of similar rates the proportion between cleavages and condensations is 50 : 50, whereas in case of particular environments or particular living systems this proportion could constitute a relatively free parameter. In neutral environments the variation of this parameter do not have particular consequences: however, this is not the case where the concentration of small species is kept constant.

---

[14] It is worth stressing that an autocatalysis is a strongly connected component, nevertheless we decided to deal with them separately.

[15] At high catalysis levels almost each RAF owns at least one SCC and one autocatalysis, so with our measurements it is not possible to observe the exact RAFs' structure. This aspect will be analyzed in further works.

Indeed, since the membrane tends to allow the free transfer to small molecules only, the apparent symmetry of this situation is broken: whereas it is possible to continuously construct long molecules by simply concatenating (by means of recursive condensation reactions) the always present small ones, it is not possible to continuously cleave long molecules, because their presence is not similarly guaranteed. Thus, it is not possible to observe RAF if just cleavages are present. Figure 5 shows the influence of different levels of the proportion between cleavages and condensations on the presence of RAF in chemical sets and on the position of their previously commented transition, in the case "BUF_2" scenario. In this case, note that, because the product of reactions involving buffered molecules are buffered molecules and because these buffered molecules do not catalyze anything, in presence of only cleavages nothing happen (data not shown here).
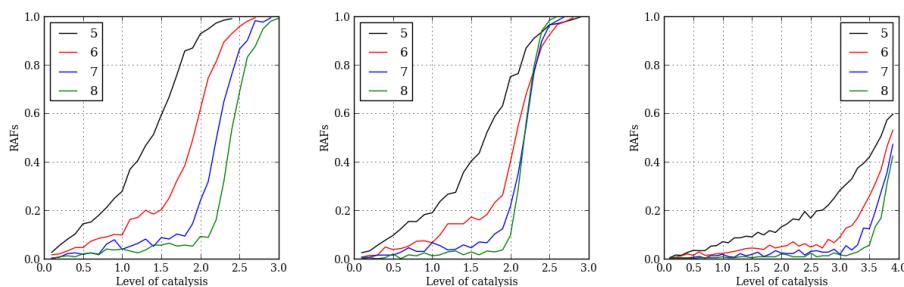

Fig. 5. The fraction of simulations with a food set composed of all the species up to length 2 showing at least 1 RAF, by varying the average level of catalysis and the maximum chemical species length. The plots show situations where the proportion between condensations and cleavages is $100 : 0$ (left), all condensations; $75 : 25$ (middle); $25 : 75$ (right). The case $0 : 100$ (all cleavages) is devoid of RAFs, hence graph is not shown here while the case $50 : 50$ is shown in figure 4 left top.

Note that if we allow all the possible condensations and cleavages among the existing molecules up to length L, the relative proportion of cleavages and condensation is obtained by a purely combinatorial argument. In this case the number of conceivable reactions is equal to:

$$\hat{R}_1 = \sum_{i=2}^{L} [2^i \cdot (i-1)] + |S|^2 \tag{2}$$

where the first term stands for the cleavage reactions and the second term for the condensation reactions [12]. It is worth stressing that, as L increases and preventing the presence of reverse reactions, the number of conceivable condensation reactions tends to become higher than the number of conceivable cleavages, according to the ratio $[\sum_{i=2}^{L} 2^i \cdot (i-1)]/|S|^2$. In other words, the "purely
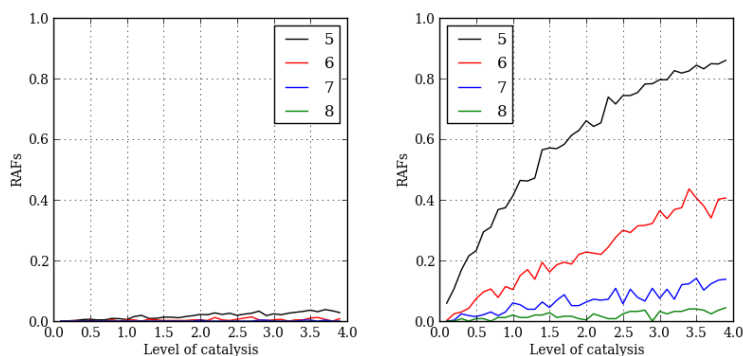
Fig. 6. Fraction of random systems, with the "purely combinatorial" quotient between cleavages and condensations, having at least 1 RAF, with food till length 2 and (c) till length 3

combinatorial" proportion between cleavages and condensations is in favor of condensations, and this advantage increases with the diversity of the present chemical species.

Following this idea, there may exist lots of condensations whose products are longer than L. This second method, applied to the same number of chemical species, leads to a very different number of reactions: so the question arises as how can we compare these two methods, and what can we say about their physical plausibility. We can observe that in (the relatively simple) pre-biotic environment very long molecules rarely appear, because of generic physical-chemical constraints. Thus, we think that the more correct model interpretation of the L limit should be: "chemical species longer than L are not suitable for further chemical processes, because of physical-chemical constraints" (for example, lack of solubility). A large fraction of the condensations leads therefore to not suitable products, which cannot play an active role within the system; thus the creation of this kind of "garbage" can inhibit the formation of RAFs (see figure 6).

4 Conclusion

The comprehension of the properties of sets of many interacting molecules poses formidable problems, and it is still a big challenge to obtain them in laboratory.

In this work we aim at revealing some properties of abstract realizations of catalytic reaction networks, with particular regard to the presence of structures potentially able to sustain their own existence and growth (the Reflexively Autocatalytic Food generated sets, briefly RAFs). Our investigations take into account different scenarios, involving different distributions of the chemicals' catalyzing capabilities, the presence or absence of backward reactions, and the

presence or absence of long chemical species that lead to the accumulation of useless chemical species.

We confirmed that there is an ample region where the systems have autocatalytic structures, which are nevertheless unable to self-sustain (RAFs are not present); interestingly, this region shows the same amplitude across different topologies and/or settings. At a sufficient level of catalysis, complete RAFs emerge, whose presence and composition (different proportion of strongly connected components, autocatalysis and linear chains) depend on some parameters (average level of catalysis, number of buffered chemical species, proportion between cleavages and condensations) and is independent on other characteristics (uniform or scale-free topology, presence or absence of backward reactions). Finally, we discussed two different ways to build artificial networks and their physical implications.

We think that these hints could be useful for the comprehension of the emergence of sets of molecules able to collectively self-replicate in OoL scenarios, and for the designing of new artificial protocells.

Acknowledgements

References

1. B. Alberts, A. Johnson, J. Lewis, M. Raff, K. Roberts, and P. Walter. Molecular Biology of the Cell, Fifth Edition. Garland Science, 4 edition, 2005.
2. R. Bagley and J. D. Farmer. Spontaneous emergence of a metabolism. Artificial Life II. Santa Fe Institute Studies in the Sciences of Complexity, X:93–141, 1992.
3. A.-L. Barabási and R. Albert. Emergence of scaling in random networks. Science, 1999.
4. T. Carletti, R. Serra, M. Villani, I. Poli, and A. Filisetti. Sufficient conditions for emergent synchronization in protocell models. J Theor Biol, 254(4):741–751, 2008.
5. R. Diester. Graph theory. Springer, Berlin Heidelberg, 2005.
6. F. Dyson. Origins of life. Cambridge University Press, 1985.
7. M. Eigen and P. Schuster. The hypercycle. a principle of natural self-organization. part A: Emergence of the hypercycle. Die Naturwissenschaften, 11(64), 1977.
8. M. Eigen and P. Schuster. The hypercycle: a principle of natural self-organization. Springer, 1979.
9. J. Farmer and S. Kauffman. Autocatalytic replication of polymers. Physica D: Nonlinear Phenomena, 220:50–67, 1986.
10. A. Filisetti, A. Graudenzi, R. Serra, M. Villani, D. De Lucrezia, R. M. Fuchslin, S. A. Kauffman, N. Packard, and I. Poli. A stochastic model of the emergence of autocatalytic cycles. Journal of Systems Chemistry, 2(1):2, 2011.
11. A. Filisetti, A. Graudenzi, R. Serra, M. Villani, D. De Lucrezia, and I. Poli. The role of energy in a stochastic model of the emergence of autocatalytic sets. In

L. T, G. M, H. Bersini, P. Bourgine, M. Dorigo, and R. Doursat, editors, <u>Advances in Artificial Life, ECAL 2011 Proceedings of the Eleventh European Conference on the Synthesis and Simulation of Living Systems</u>, pages 227–234. MIT Press, Cambridge, MA, 2011.

12. A. Filisetti, A. Graudenzi, R. Serra, M. Villani, R. M. Füchslin, N. Packard, S. a. Kauffman, and I. Poli. A stochastic model of autocatalytic reaction networks. <u>Theory in biosciences = Theorie in den Biowissenschaften</u>, pages 1–9, Oct. 2011.

13. A. Filisetti, R. Serra, T. Carletti, M. Villani, and I. Poli. Non-linear protocell models: synchronization and chaos. <u>The European Physical Journal B</u>, 77(2):249–256, June 2010.

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

15. T. Ganti. Chemoton Theory, Vol. I: Theory of fluyd machineries: Vol. II: Theory of livin system. <u>New York: Kluwer Academic/Plenum</u>, 2003.

16. W. Hordijk, S. A. Kauffman, and M. Steel. Required Levels of Catalysis for Emergence of Autocatalytic Sets in Models of Chemical Reaction Systems. <u>International Journal of Molecular Sciences</u>, 12(5):3085–3101, May 2011.

17. W. Hordijk and M. Steel. Detecting autocatalytic, self-sustaining sets in chemical reaction systems. <u>Journal of theoretical biology</u>, 227(4):451–461, Apr. 2004.

18. W. Hordijk, M. Steel, and S. Kauffman. The Structure of Autocatalytic Sets: Evolvability, Enablement, and Emergence. page 12, May 2012.

19. S. Jain and S. Krishna. Autocatalytic set and the growth of complexity in an evolutionary model. <u>Phys Rev Lett</u>, 81:5684–5687, 1998.

20. S. Jain and S. Krishna. A model for the emergence of cooperation, interdependence, and structure in evolving networks. <u>Proc. Natl. Acad. Sci., USA</u>, 98(2):543–547, 2001.

21. S. A. Kauffman. Autocatalytic sets of proteins. <u>J Theor Biol</u>, 119(1):1–24, 1986.

22. S. A. Kauffman. <u>The origins of order</u>. New York: Oxford University Press, 1993.

23. P. Luisi, F. Ferri, and P. Stano. Approaches to semi-synthetic minimal cells: a review. <u>Naturwissenschaften</u>, 93(1):1–13, 2006.

24. S. S. Mansy, J. P. Schrum, M. Krishnamurthy, S. Tobé, D. A. Treco, and J. W. Szostak. Template-directed synthesis of a genetic polymer in a model protocell. <u>Nature</u>, 454(7200):122–125, 2008.

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

26. S. Rasmussen, L. Chen, M. Nilsson, and S. Abe. Bridging nonliving and living matter. <u>Artificial life</u>, 9(3):269–316, Jan. 2003.

27. T. Rocheleau, S. Rasmussen, P. E. Nielsen, M. N. Jacobi, and H. Ziock. Emergence of protocellular growth laws. <u>Philos Trans R Soc Lond B Biol Sci</u>, 362(1486):1841–1845, 2007.

28. L. Schmidt. <u>The Engineering of Chemical Reactions</u>. Oxford University Press, 1998.

29. D. Segre, D. Lancet, O. Kedem, and Y. Pilpel. Graded autocatalysis replication domain (gard): kinetic analysis of self-replication in mutually catalytic sets. <u>Orig. Life Evol. Biosph.</u>, 1998.

30. R. Serra, T. Carletti, and I. Poli. Syncronization phenomena in surface-reaction models of protocells. <u>Artificial life</u>, 123(2):123–38, 2007.

31. R. Serra, A. Filisetti, M. Villani, A. Graudenzi, C. Damiani, and T. Panini. A stochastic model of catalytic reaction networks in protocells. Under Revision, 2014.

32. R. R. V. Solé, A. Munteanu, C. Rodriguez-Caso, J. Macía, and J. Mac\'\ia. Synthetic protocell biology: from reproduction to computation. Philos Trans R Soc Lond B Biol Sci, 362(1486):1727–1739, 2007.

33. P. F. Stadler. Dynamics of autocatalytic reaction networks. IV: Inhomogeneous replicator networks. Bio Systems, 26(1):1–19, Jan. 1991.

34. J. W. Szostak, D. P. Bartel, and P. L. Luisi. Synthesizing life. Nature, 409(6818):387–390, 2001.